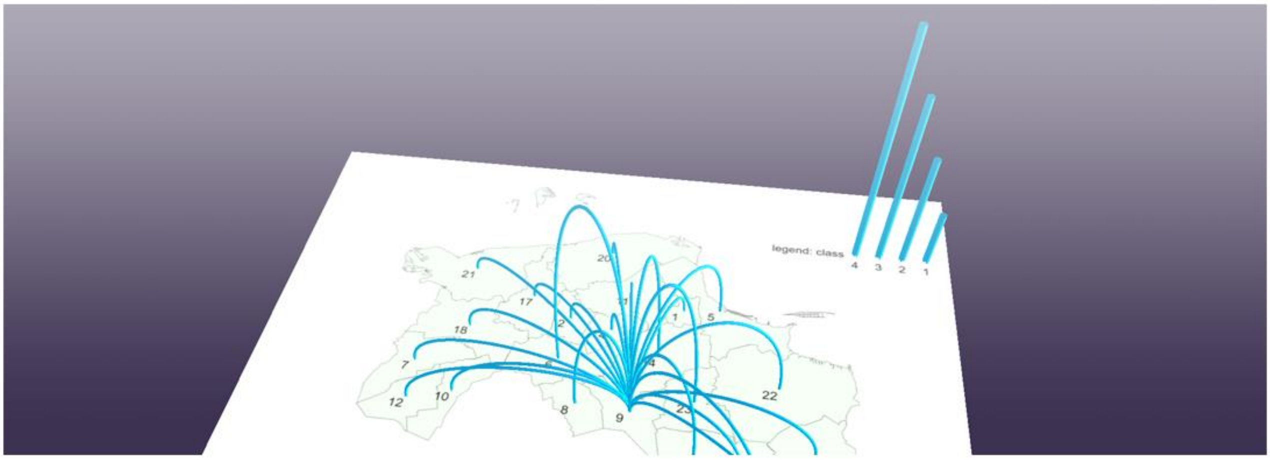
GIMA

Geographical Information Management and Applications



Reducing visual clutter on dynamic 3D flow maps

MSc. thesis

student	Niels Struis
email	n.struis@students.uu.nl
solis ID	4115732
supervisor	Menno-Jan Kraak
responsible professor	Raul Zurita-Milla

date	29 - 10 - 2019

Universiteit Utrecht

TU Delft

UNIVERSITY OF TWENTE.

WAGENINGEN UR

The picture on the front page of this report is a screenshot of one of the flow maps used in the usability test.

Preamble

This thesis is one of the final steps in the fulfilment of my master program *Geographical Information Management and Applications*. Last year, I had the chance to learn more about cartography and product design while creating, testing and performing a usability test for flow maps. I gained personal insights in my learning and working abilities as well. First of all, I would like to thank my supervisor, Professor Menno-Jan Kraak, for introducing me to the topic, providing me with valuable feedback and sharing his enthusiasm and expertise with me. This research could have not been realised without the collaboration with PhD candidate Yuhang Gu. His technical expertise contributed to the development of 3D dynamic flow maps for the usability test. Our brainstorm sessions provided a valuable guidance for the construction of this report. Also, I am grateful for the feedback on the quality of the usability test received from professor Corné van Elzakker and Ieva Dobraja during the focus group session. The motivation and efforts provided by the participants is highly appreciated. At last, I am thankful for the motivational support of my GIMA-peers, friends and family along the way.

Summary

Because of the increasing availability of geographical data, there is a growing need for maps enriched with a larger amount of information. However, the increasing amount of data visualized in one map can weaken the map readability. An example of this phenomenon is visual clutter, which can occur on flow maps. Visual clutter is defined as "*a limitation of the map readability caused by intersection or clustering of multiple flows*" (Jenny et al., 2018). This obstruction of the map objects negatively influences the map usability. The available measures for reducing visual clutter are only applicable to flow maps represented in 2D. Therefore, a usability test is performed on flow maps in a 3D display environment by using both a PC screen and an iPad in combination with the mixed reality (MR) technique. The following main research question is used:

> *How can visual clutter on flow maps be reduced and to which extent is the usability of a dynamic 3D flow map influenced when visualized on a PC screen compared to an iPad screen in a mixed reality environment by using two different flow line symbologies?*

A qualitative method is used for its explorative and descriptive research capabilities. The term usability as defined by the ISO standard is the measurement tool for indicating a potential reduction of visual clutter by comparing the results on the efficiency, effectiveness and satisfaction (International Organization for Standardization, 2018). Additionally, the participants are asked to speak out their reasoning behaviour by using a think aloud method which is traced by written notes and audio recording. Additional recordings are used as back up material: video recording and screen recording. The combination of these information sources are qualitatively analysed by comparing its final results.

At the finalizing stage of the usability test design, a focus group discussion with domain experts is performed for collecting feedback for enhancing the usability test, followed by pilot tests. For the usability test, the two main topics to be covered are based on the four map stimuli; two flow line symbologies and the two display environments. For visualizing quantitative attribute information, a variation in either the arc height or the stroke width is applied. Both a 3D map on a screen as well as a 3D MR map are used as the display environments. The test consists of a familiarization phase for training the participants with the hardware and software, the main test as a questionnaire in correspondence to the flow maps and a satisfaction interview at the final stage of the test.

The collected questionnaire data is analysed with descriptive statistics, using boxplot figures in combination with the comments made by the participants during the tests. Based on these results, it is concluded that for the efficiency, effectiveness and satisfaction the usability of the maps displayed in the PC environment is better than for the MR maps, and the usability of the maps with a flow volume symbology is better than the arc height. Although MR can enable new visualization opportunities, the usability of a flow map does not automatically improve when compared to the same 3D map displayed on a PC screen.

List of abbreviations, figures and tables

Abbreviations

AI	Artificial Intelligence
AR	Augmented Reality
GIS	Geographical Information System
GVis	geographic visualization
MR	Mixed Reality
OD	Origin destination
UCD	User centred design
VR	Virtual Reality

the four map stimuli	the display environment on the PC, the display environment on augmented reality, the flow line symbology arc height and the flow line symbology line volume
the three usability characteristics	efficiency, effectiveness and satisfaction

Figures

	Figure	page number
1.1	Screen capture of the Wind map and weather forecast service	14
1.2	The 50 largest migration flows between U.S. states	14
1.3	Maps with curved and straight flow lines	15
1.4	YouTube video; Mapping the World's Immigration Flows Country-by-Country, as an example of an animated flow map	15
2.1	Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812–1813	19
2.2	Visual variables	21
2.3	Cartography cubed	22
2.4	The scope of visual analytics	23
2.5	Temporal legends	24
2.6	Visualization options for the directionality and progress of time	25
2.7	Examples of flow line symbology of flow lines in the third dimension	26
2.8	Miligram's reality-virtuality continuum	27
2.9	User-centered design process	28
3.1	The four main phases of user centered design approach	32

3.2	Scheme of the usability test workflow used for this research	32
3.3	Example of map with centroids for each municipality within a province	36
3.4	Flow map concept in a 3D environment. The map is perceived on the screen	37
3.5	Flow map concept in a 3D AR environment. The map is perceived on a flat surface such as a table.	37
3.6	The two flow line symbologies used in this research	38
3.7	Usability test map TU1	43
3.8	Usability test questionnaire, map TU1	44
3.9	Focus group topic list	45
5.1	Boxplot of the time duration of all maps, the divided in four groups	59
5.2	Boxplot of scores of each map, divided in four groups	60
5.3	Boxplot of the satisfaction scores of eight of the total maps, divided in four groups	61
5.4a	Users experience with maps, 3D maps and AR environment.	62
5.4b	Sub groups based on participants' experience with 3D and AR	62
5.5	Average time spend on maps in AR and screen	63
5.6a	Time spent on each question from participants with AR experience and no AR experience	63
5.6b	Time spent on each question from participants with 3D experience and no 3D experience	63
5.7	The mean scores for one map grouped by map reading experience	64
5.8a	Time spent on each question from participants with AR experience and no AR experience	64
5.8b	Time spent on each question from participants with 3D experience and no 3D experience	64

Tables

Table		page number
2.1	Overview of all visualization techniques for reducing visual clutter in 2D flow maps	26
3.1	An overview of all contributions to this thesis research by Yuhang Gu	34
3.2	Navigation instructions for the maps in both display environments	39
3.3	The usability characteristics per flow line symbology applied in the usability test	39
3.4	The usability characteristics per display environment applied in the usability test	39
3.5	Overview of the flow map characteristics used in this research	40
3.6	Overview of all maps used for the familiarization and main usability test, divided in main- and subgroups	41

3.7	Sequencing of the maps per user ID	42
3.8	An OD matrix with the level two connections marked in red	43
3.9	Likert scale answer options	48
4.1	Overview of the available recordings per participant ID	52
5.1	Pros and cons of the display environments, based on the participant’s comments	56
5.2	Pros and cons of two display environments, based on the participant’s comments	58

Contents

Preamble 3

Summary 5

List of abbreviations, figures and tables 6

Contents 9

1. Introduction 13

1.1 Problem statement 14
1.2 Research questions 17
1.2.1 Main research question 17
1.2.2 Sub research questions 17
1.2.3 Hypothesis 17
1.3 Scope, limitations and acknowledgements 17
1.4 Research goal 18
1.5 Relevance 18

2. Literature study 19

2.1 Introduction chapter 19
2.2 Flow map characteristics 19
2.3 Design principles in cartography 20
2.3.1 Basic elements of visualization 20
2.3.1.1 Visual variables and mapping environments 21
2.3.1.2 Interaction and communication 21
2.3.1.3 Visual analytics 22
2.3.2 Design principles for flow maps 23
2.3.2.1 Variables 23
2.3.2.2 Spatio-temporal design 23
2.3.3 Available solutions for minimizing visual clutter 26
2.4 The development of modern techniques in cartography 27
2.5 Usability study 28
2.5.1 Usability and readability 29

3. Methodology - design 31

3.1 Introduction chapter 31
3.2 Research method 31
3.2.1 Qualitative research method 31
3.2.2 User centred design 31
3.2.3 Usability test workflow 32
3.2.4 User characteristics 34
3.2.5 Research collaboration 34
3.3 Map design 35
3.3.1 General 35

3.3.1.1 Data preparation 35
3.3.1.2 Map features 35
3.3.2 Display environments 36
3.3.3 Flow line symbology 37
3.3.4 Map navigation 39
3.4 Questionnaire design 40
3.4.1 Stimuli sequencing 40
3.4.2 Questions 43
3.4.2.1 Difficulty levels 43
3.4.2.2 Scoring method 44
3.5 Focus group 45
3.5.1 Preparation 45
3.5.2 Results 46
3.6 Usability test script 47
3.6.1 Introduction 47
3.6.2 Familiarization 47
3.6.3 Main test 47
3.6.4 Satisfaction interview 48
3.7 Final preparations and challenges 49
3.7.1 Pilot testing 49
3.7.2 Facilities and invitations 49
3.7.3 Materials 49
3.7.4 Challenges 50
4. Methodology - performance 51
4.1 Introduction chapter 51
4.2 Conditions 51
4.2.1 Facilities 51
4.2.2 Participants 51
4.2.3 Material 52
4.2.4 Questionnaire design 53
5. Results 55
5.1 Introduction chapter 55
5.2 Participant observation 55
5.2.1 Display environments 55
5.2.2 Flow line symbologies 57
5.3 Participant performance 59
5.3.1 Efficiency 59
5.3.2 Effectiveness 60
5.3.3 Satisfaction 61
5.4 Participant experience 62
5.4.1 Efficiency based on experience 62
5.4.2 Effectiveness based on experience 64

6. Conclusion 65

7. Discussion 66

7.1 Interpretation 66

7.2 Limitations 67

7.3 Recommendations 67

8. References 69

A1. Usability test questionnaire 75

A2. Manual usability test 93

A3. Targets MR 95

A4. Focus group transcript 99

A5. Notes of all sessions 107

1. Introduction

The increasing data accessibility, computation power, storage space and the smaller sizes of computational devices is stimulating the growing map usage, which has started a '*Renaissance in cartography*' (Tsou, 2011). Technological developments, especially in web cartography, have resulted in the emergence of interactive and dynamic visualizations compared to the traditional static printed maps. An example of a map which is being affected by this development is the flow map: a representation of linear movements of both tangible or intangible objects varying in quantity through space between an origin and destination location (Field & Saunder, 2018; Börner, 2015). However, when the number of displayed flow lines on a map is increasing, the map becomes less readable because of the overlapping and intersection of the line segments. For this thesis, alternative visualization approaches for flow maps are tested on its usability for reducing this readability effect called visual clutter (Gu, Kraak, & Engelhardt, 2017).

Cartography can be considered as the combination of art and the science for the presentation of information on a map (Kraak & Ormeling, 2013). A map is regarded as a graphical abstraction of reality. Therefore, careful considerations have been made to decide what information to present, for creating a clear and accurate map. Additionally, a simplification of information implies that a map can never be fully objective, but cartographic rules are a guide to make it as objective as possible (Kraak, 2011). Traditionally, cartographers were responsible for the map readability. However, only a limited share of the total maps are being controlled by these experts today (Harrie, Stigmar, & Djordjevic, 2015). Cartography has changed from a supply- to a demand driven process in which map making has become more accessible for people without cartographic expertise. However, "*it should remain the task of the spatial data handling world to test the maps on its effectivenes*s" (Kraak, 1998).

The adaptation of new visualization techniques in cartography is slow in contradiction to other industries such as video production. As an example, flow line visualizations are displayed in a three dimensional (3D) presentation of areas in the Netherlands in the Dutch documentary series "*Nederland van boven*" (VPRO, 2018). Although these visualizations are used for story telling for a large audience, the underlying scientific cartography techniques for flow maps in particular have not been researched yet.

Meanwhile, data availability and dataset sizes are increasing, partly because of the expanding usage of sensors for the Internet of Things (Townsend, 2013). Large datasets require suitable analysis and visualization techniques, because the simplified presentation of highly detailed information needs more pre-processing power in order to achieve a readable map (Stephen & Jenny, 2017). The amount of presented information and the spatial distribution are the two most important measures for map readability (Harrie et al., 2015). The increasing wealth of information could be a challenge for map making, because more initial selection and analysis is needed. As stated by Monmonier, "*to avoid hiding critical information in a fog of detail, the map must offer a selective, incomplete view of reality*" (2005). This illustrates how maps are never a complete representation of reality, but a selection which can easily be comprehended by the reader.

Figure 1.1: Screen capture of the Wind map and weather forecast service. (Lukačovič, 2018)

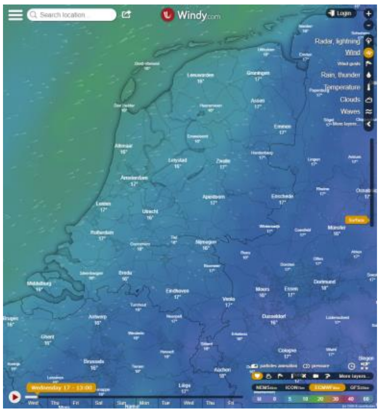
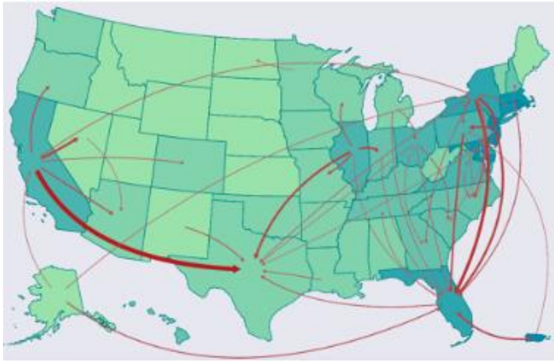


Figure 1.2: The 50 largest migration flows between U.S. states (Stephen & Jenny, 2017)



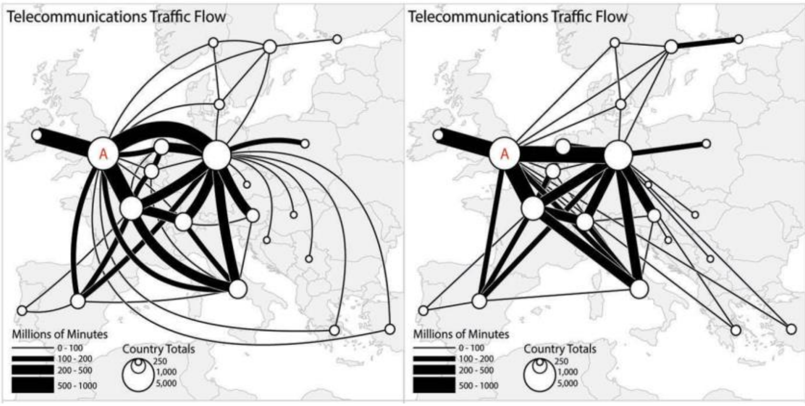
The flow line symbology of dynamic movements of objects through space and time emerged during the early 1800s, but is being used more because of the increasing data availability in the last few years (Claudel, Nagel, & Ratti, 2016). The characteristics of the flow's origin, destination, direction and speed are attributes which could be displayed simultaneously by using a spatial or time component (Gu et al., 2017; Jenny et al., 2016). The attribute representation is depending on the stylistic choices determined by the map designer. For earth scientists, flow maps are used for displaying physical dynamics, such as a meteorological wind map (figure 1.1), but it can be used to illustrate societal developments as well. An example of a societal development displayed on a dynamic flow map is the map of migration patterns in the United States (Stephen & Jenny, 2017) (figure 1.2).

1.1 Problem statement

Unlike flow diagrams, the beginning and end nodes on a flow map are predetermined because these are geographically bound. Therefore, when the amount of presented flows is increasing, the there is a higher probability for visual clutter to occur (Gu et al., 2017). Visual clutter is defined as "*...a limitation of the map readability caused by intersection or clustering of multiple flows*" (Jenny, et.al., 2016). Design alternatives have been created to minimize visual clutter, but these are not effective for more information dense maps (Jenny et.at., 2017). Although a large amount of presented data does not necessarily result in visual clutter, the chances are higher because the more flow lines are presented, the more difficult it is to arrange an spatial distribution with an effective readability (Harrie et al., 2015). The ultimate goal is to create an easily readable map with a high communication quality, while it should contain complex and detailed information (Dibiase, Maceachren, Krygier, & Reeves, 1992). However, an oversimplified visualization should be prevented because losing too much information could cause a different interpretation of the map data.

Potential solutions for visual clutter are alternative cartographic variables (Fisher, Dykes, & Wood, 1993). An alternative map attribute design option is illustrated in figure 1.3. Encoded line styles can be applied to represent the attribute information, such as colour, line volume, stroke pattern, curvature or transparency (Romat et al., 2018). As an example, curved flow lines have a better readability compared to straight lines for 2D flow maps, because the line segments are more evenly spatially distributed (figure 3; Jenny et al., 2016). As of today, these techniques are mainly designed for two dimensional maps. Therefore, limited literature is available on displaying flow maps in 3D , although visualization in the third dimension would allow for more design options (Gu et al., 2017).

Figure 1.3: Maps with curved and straight flow lines, Jenny et.al. 2016



Changing the display environment is another potential solution for reducing visual clutter, since the usability has not been investigated for dynamic flow maps in a three dimensional view. Static 2D maps are a representation of the area as seen from above. Dynamic maps can be displayed in either 2.5D or 3D. The 2.5D representation is defined as the display of 3D data using a 2D medium (i.e. a map printed on a piece of paper) without an immersive or interactive capability (Read, Phillipson, Serrano-Pedraza, David Milner, & Parker, 2010). An example of a 2.5D representation is a visualization of relief data such as a digital elevation model (DEM), which is only representing one elevation (z) for a given (x, y) position. While the 2.5D environment enables a descriptive perception of reality, the 3D environment is an immersive perception which allows the user to observe the map from any desired perspective (Cambray & Yeh, 1994). Examples of a 3D representation are a stereoscopic image or augmented reality (AR).

Animation allows to display a variation of speed for the movement of individual dots per flow line, more detailed attribute information can be presented without causing visual clutter in the map frame (Gu et al., 2017). An example of animation in cartography is the YouTube video (see the screen capture in figure 1.4). The user is able to recognize and track changes of geographic development when watching an animated flow map video. Potentially, new observations could be made by the end user when viewing the map, which indicates the shift to the demand driven map making process (Harrower & Fabrikant, 2008). In addition to the stationary design elements, animated visual encoding can be applied as well. Romat (et al. 2018) introduced animated edge textures and the speed, pattern and frequency of individual particles as the visual characteristics for this mapping type.

Figure 1.4: Screen capture of the YouTube video; Mapping the World's Immigration Flows Country-by-Country, as an example of an animated flow map. Metrocosm, 2016

There is an increasing need to effectively visualize information dense maps, and the there is little scientific literature available on the alternative techniques to reduce visual clutter on flow maps. Both are indicators of the urgency of this matter. Clarification is needed on the best practices of visualizing flow maps with the currently available techniques for improving the its readability and reducing visual clutter. Augmented reality is a rapidly developing technique which can be used for 3D data visualization. Using AR allows for a rich and interactive experience, and it is highly available to the public because of the integration on the latest released smartphones.

When using AR, it is not needed for the user to move around the real world for viewing the data content. Therefore, the ability to display a 3D model on a flat surface (i.e. a floor or table) could be sufficient to solve the visual clutter problem. In this case, the map can be presented on a mobile device while the camera is aimed at a flat surface. Augmented reality is defined as *"an integration and merging of the real and virtual worlds in which physical and virtual objects complement and interact with each other."* (Kim, Hwang, & Zo, 2016). However, the virtual map objects are only stitched to a flat table surface in the real world, instead of a more immersive experience where the user's geographical location is influencing the experience as well. Therefore, the term mixed reality (MR) is used for this research.

Mixed reality has not been applied to a geographic information system (GIS) in combination with flow maps yet, according to the available literature. However, MR is a hot research topic in many other fields, i.e. the tourism industry for enhancing visitor experiences and for improving medical equipment (Tussyadiah, Jung, & tom Dieck, 2018; Moro, Štromberga, Raikos, & Stirling, 2017). Because the user can interactively view a 3D object from every desired perspective by moving the device around, a complex whirl of lines displayed in a 3D flow map might be easier to identify. Therefore, applying MR to a flow map visualization could improve the readability and usability.

1.2 Research questions

The scope of the research is defined with the aid of a main research question and three sub research questions:

1.2.1 Main research question

> *How can visual clutter on flow maps be reduced and to which extent is the usability of a dynamic 3D flow map influenced when visualized on a PC screen compared to an iPad screen in an augmented reality environment by using two different flow line symbologies?*

1.2.2 Sub research questions

1. What are the characteristics of a dynamic 3D flow map with visual clutter?
2. To which extent is the usability of dynamic 3D flow maps influenced by the four map stimuli (two display environments and two flow line symbologies) based on the results of the usability test?
3. To which extent can visual clutter on dynamic 3D flow maps be reduced?

1.2.3 Hypothesis

The following statement is used as the hypothesis of this research:

Visual clutter is reduced when flow maps are visualized in the MR display environment when compared to the PC environment.

1.3 Scope, limitations and acknowledgements

- Only non-branching origin-destination (OD) flow map type is used in this research, for which maps are presented on top of a geographical map layer (Jenny et.al, 2016). This way, both the flow information with reference to the movement data as well as a sense of space is achieved.

- The migration data between municipalities in 2017 is used for the maps in the usability test for the provinces of Utrecht, Friesland and Groningen (CBS, 2017).

- For testing the usability of the flow maps, four different map stimuli are investigated in this research. Two display 3D display environments are used: a PC screen and an iPad for using MR. Additionally, two flow line symbologies are used, the variation in line volume and arc height for the flow lines. These stimuli are abbreviated as: MR, PC, volume and height. This limited set of four variables is necessary for avoiding complications for analysing the results afterwards, since a qualitative research method is applied.

- Mixed reality is chosen as the technique for an alternative display environment to the PC screen because of its good accessibility in technology and the lack of its application to cartographic research today. Therefore similar techniques such as virtual reality or stereoscopy are not considered in this research.

1.4 Research goal

The goal of the usability test is to gain knowledge on the usability of alternative visualization methods for flow maps for reducing visual clutter. By testing the usability of new visualization techniques, more insight is generated on the visualization requirements specifically tailored to a 3D visualization on a PC screen and augmented reality on an iPad. Therefore, the information on the usability of these new techniques can be used to further improve the visualization of a flow map, for future equality improvement with a good communication quality which should contain complex, rich and detailed information.

1.5 Relevance

Two decades ago, is has been stated that there is an opportunity for geographers and cartographers to benefit from the new technological possibilities for visualization research (Fisher et al., 1993). However, the design of flow maps is mainly based on expert intuition and aesthetic considerations today (Jenny et al., 2018). According to (Andrienko et al., 2010), "*researchers should find approaches to deal with the complexities of the current data and problems and find ways to make analytical tools accessible and usable for the broad community of potential users to support spatiotemporal thinking and contribute to solving large range of problems.*" The rising availability of spatial and temporal data creates both new possibilities and problems for data visualization. A number of small design alternatives have been created to prevent visual clutter, but this problem appears to remain, especially with the increasing data availability and the desire to create maps with a higher information density (Jenny et al., 2018). Therefore, there is a strong need for an effective spatiotemporal visualization tool for flow maps (Claudel et al., 2016).

Displaying a flow map in 3D could improve the readability because the user can be enabled to view the flow lines from any perspective, therefore in theoretical hypothesis, individual flow lines could be better observed. Although similar graphics of flows in a 3D environment have widely been used in other industries, i.e. by graphic designers in the film industry, there is hardly any cartographic research about this new visualization method (VPRO, 2018). Yang is stating that the technology may replace traditional flat panel displays in many situations, because the use of 3D could resolve visual clutter for complex flow maps using virtual- and augmented reality (Yang et al., 2018). The academic relevance of this thesis research is to fill the gap in the available literature on what cartographic rules and techniques are most suitable for a reduced visual clutter in flow maps.

2. Literature study

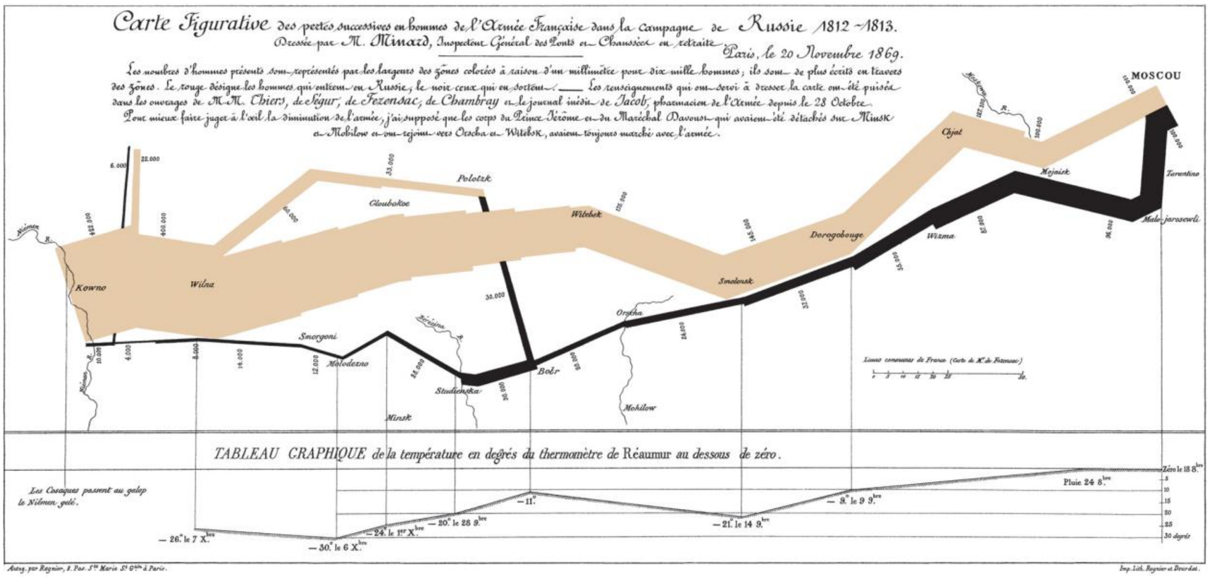
2.1 Introduction chapter

In this chapter, flow maps and visual clutter as its related ‘problem’ are explained, along with practical examples. Afterwards, the cartographic visualization elements are compared to the currently known design principles for flow maps and existing solutions for visual clutter. This chapter ends with information on usability testing, since the research data is collected with this method.

2.2 Flow map characteristics

Flow maps are one of many existing map types, which belongs to the group of line maps as the display of individual linear movements or a network (Börner, 2015). Flow maps have usually been used to present origin-destination (OD) data (Gu et al., 2017). Linear movements are symbolized by lines on a map, to include both the spatial trajectory as well as additional information by the applied line styling variations. The OD data can be stored in a matrix, but the added value of the presentation on the map is the ‘sense of space’ which is achieved when the lines are displayed on top of a geographical base map layer. The nodes (point features on a map) representing the origin and destination are bound to a geographical location (Jenny et al., 2018). The figurative map by Minard is a leading cartographic example of the earliest flow maps, originating from 1869 (figure 2.1). This map is famous as the best statistical graphic ever drawn, because of its ability to present a combination of six different sets of data; the rivers, cities and battles represent the geography, the route, the direction, the number of soldiers remaining (the path is becoming smaller; one millimetre represents 10.000 men), temperature and time (at the bottom of the figure).

Figure 2.1: Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812–1813. Drawn by Charles J. Minard, 1869. Field, 2018



Flow lines can represent a multitude of data attributes; the flow origin, destination, density, direction, type and speed (Gu et al., 2017; Jenny et al., 2018). These attributes can contain either a quantitative or qualitative data type. The line trajectories can represent either tangible (people or goods) or intangible (ideas, expertise or digital documents) objects (Börner, 2015). Each flow line is capable to display a multitude of attributes simultaneously. In contrast to flow charts, the beginning and end nodes of each flow trajectory are inseparably bound to a geographical location. Depending on the scale level used for the map, these locations does not have to be exact but are used to depict the origin and destination in the geographical context area. Therefore, when flow lines are displayed on a map, directly layered on top of the base map, there are limited options for rerouting the lines or nodes to reduce visual clutter compared to corresponding flow charts where there is no sense of space (Jenny et al., 2018).

2.3 Design principles in cartography

2.3.1 Basic elements of visualization

Cartography is the science and art of making maps as a representation of the real world. Maps are a representation of reality, which are created by a carefully selected set of criteria. A questions cartographers should be able to answer for themselves when creating a map is "*How to say what to whom and is it effective?*" (Kraak, 1998). This questions implies that both the method in which the content is displayed (how), the selected map content (what), the target audience (whom) and the usability of the information transfer (effective) have been considered in the map making process. The visualization process can be applied to all types of data for the production of a map; First, (raw) data is analysed to convert it to information. Only the relevant entries are selected by filtering/querying, the map extent zoom level is chosen, and potential further analysis is performed. The final desired details are implemented, such as the map styling and layout, for completion of the final map (Keim, Mansmann, Schneidewind, Thomas, & Ziegler, 2008).

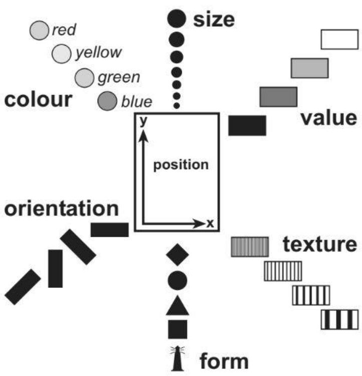
The development of new user friendly technologies enabled people both to access and create map content more easily, mainly because of the Internet as a marketplace of ideas, knowledge and social change (Börner, 2015). The increasing accessibility to data sources and the tools to create maps, including open source data and software, are enabling everyone with the access of a computer connected to the Internet to display and/or modify the map datasets. This liberal shift in data accessibility is a positive development considering map production costs; when organizations are collaboratively sharing their information, there is no need to 'reinvent the wheel'. However, not everyone possesses cartographic knowledge, which is needed for maintaining a high map quality. Therefore, the development of cartographic rules for the new technologies is essential for quality purposes.

2.3.1.1 Visual variables and mapping environments

Visual variables have been an essential component of map making since predefined styling characteristics of variables would allow for a more effective and efficient map, which contributes to the overall readability. Map effectiveness is considered as the accuracy for information transfer, map efficiency as the needed speed by the user to understand the information (Garlandini & Fabrikant, 2009). The seven initial visual variables were introduced by the French cartographer Jacques Bertin in ‘S*emiologie graphique*’ (1967), which have been developed over time by cartographers. For instance, Roth described an extended list of twelve visual variables: location, size, shape, orientation, colour hue, colour value, colour saturation, texture, arrangement, crispness, resolution and transparency. Since the variables are essential elements of a map visualization, these elements are continuously being redesigned and improved (Roth, 2017). The redesign of cartographic principles is an indicator that these rules are the result of critical analysis and cultural constructs.

Most variables can be applied to flow maps, however flow lines are expected to have a vector format and therefore the variables crispness and resolution would not be applicable in this situation. The cartographic decision model is a guide for the styling choice of variables, since these are depending on the data type (M. J. Kraak & Ormeling, 2013). Garlandini & Fabrikant concluded that the visual variable ‘size’ is the fastest and most accurate to read, and therefore the most efficient and effective variable (2009). However, there are no cartographic guidelines available yet for visualizing flow maps in a 3D environment.

Figure 2.2: Visual variables (van Elzakker, 2004)



Visual variables	Perception properties			
	Associative	Selective	Ordered	Quantitative
Position	+	-	-	-
Form	+	-	-	-
Orientation	+	o	-	-
Colour	+	++	-	-
Texture	o	+	o	-
Value	-	+	++	-
Size	-	+	+	++

Figure 2.5 The perception properties of the visual variables

++ = *very strong* / + = *strong* / **o** = *moderate* / - = *no*

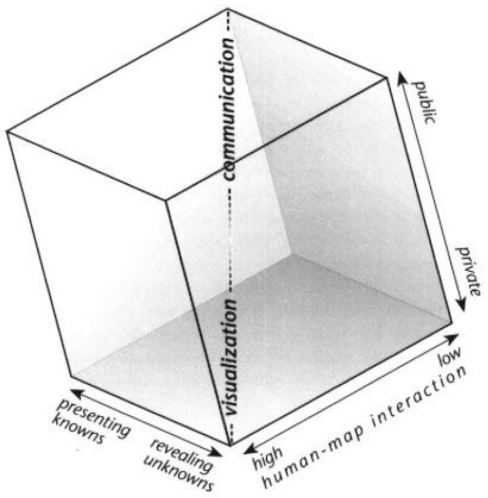
2.3.1.2 Interaction and communication

In the earlier days of cartography, the map was considered the end result. Subsequently, very little attention was paid towards the requirements of the user. Today, the main goal of cartography is not the presentation of a final map product, but the communication of information through the map as a medium (Crampton, 2001). The usage of interactive maps is increasing because of the arrival of technologies that are enabling cartographers to create more immersive map experiences. The cartography cube explains the representational value of space, which is caused by three main elements displayed on the three axis: the level of human interaction, the level of simplification and the target audience; public or private (figure 7; (MacEachren, 1994). *As stated by Crampton* (2001)*, “Traditional cartography has emphasized public use, low interactivity and revealing knowns, while visualization emphasizes private use, high interactivity and exploring unknowns”*.

When flow maps are considered using this cube, it becomes clear that the visualization quality is dependent on the method used. All methods have advantages and disadvantages, but the goal of improving the flow map visualization is to have data rich maps while reducing visual clutter. According to MacEachren, visualization is a new discipline in cartography, mostly taking place in a GIS, which can be used for data exploration since raw datasets sizes are increasing (1994).

Data interaction and quick information transfer is important for a good visualization, which "*is a way of engaging with data and deriving knowledge*" (Claudel et al., 2016). According to Koussoulakou & Kraak (1992), "*...communication capabilities of any form of visualization depend on the level of interaction*". Therefore, the addition of an interactive element is beneficial for the visualization quality of flow maps, since this allows the user to explore the underlying data more immersively.

Figure 2.3: Cartography cubed. MacEachren, 1994.

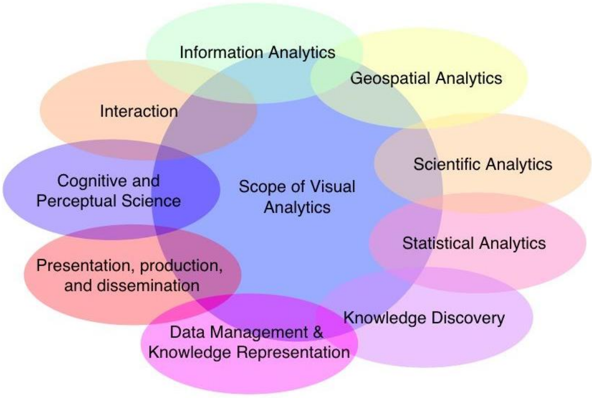


2.3.1.3 Visual analytics

The term visualization is understood as "a graphical representation of data or concepts"; a mental image of the mapped reality (Keim et al., 2008). Visual analytics is a scientific approach for enabling the examination of large datasets by decision makers by using computation techniques. As defined by Keim (2008): "*Visual analytics combines automated analysis techniques with interactive visualisations for an effective understanding, reasoning and decision making on the basis of very large and complex data sets*". Analytical processes are combined with interactive visual representations to improve the performance of complex processing, as a reaction to the development of the increasing amount and complexity of available data (Andrienko et al., 2010; Cook & Thomas, 2005). When methods are developed to make analytical tools and visualizations better accessible and usable to a diverse audience, spatio-temporal analysis processes could be more accessible. Visual analytics can improve data exploration by developing tools to "*detect the expected and discover the unexpected*" (Cook & Thomas , 2005).

Figure 2.4 illustrates the scope of visual analytics. This collaborative emphasis of visual analytics adds to the development of a greater user involvement and the democratization of knowledge and participation (Börner, 2015). There is much room for improvement on the analysis and visualization of temporal data in GIS science (Andrienko, et.al. 2010). Flow maps, as a medium for spatio-temporal data visualization, can be presented in a collaborative context such as a web application so it can be supported by a range of amateur and professional analysts. More on the tools and techniques for visualizing flow maps is discussed later in this chapter.

Figure 2.4: The scope of visual analytics, Keim et.al., 2006



2.3.2 Design principles for flow maps

2.3.2.1 Variables

Although multiple versions of visual variables are used in academic literature, only the following four variables are used for the flow map design in this research: colour, thickness, height and direction (Gu et al., 2017). Other than improving the design and aesthetic quality of the map visualization, these variables are used for the encoding of information (Roth, 2017). The application of these map types depends on the metadata type, according to the design principles in cartography. As an example, linear flow maps can only be applied to absolute ratio data (Rød, Ormeling, & van Elzakker, 2001).

2.3.2.2 Spatio-temporal design

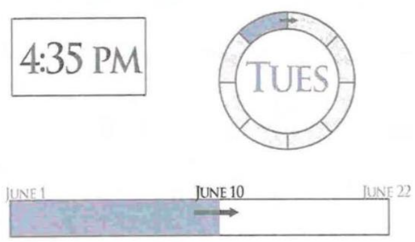
A flow map is a type of spatio-temporal data; a representation of a development through space and time. In this section is discussed how to visualize time with the available cartographic principles. The map by Charles Minard is an example of time flattening, where in this case all time slices are merged into a single 2D image (figure 2.1) (Bach, Dragicevic, Archambault, Hurter, & Carpendale, 2017). This map can be regarded as a source of inspiration for developing new visualization techniques, especially for similar flow maps on societal developments.

Hägerstrand's transdisciplinary perspective of human behaviour in space and time is known as the space-time path (2004). With this theory, the events that occur in a time fragment of an individual’s life are displayed as a vertical trajectory, in a three dimensional visualization. Within the given time fragment, the most significant steps or events are graphically presented as a point. These points act as the stations where groups can meet up or dissolve, relating to their personal paths (Pred, 1977). Hägerstrand included three constraints which could either directly or indirectly limit the individual’s freedom to move through space and time: capability-, coupling- and authority constraints (Pred, 1977).

Capability constraints are the limitations on activity caused by biological structure or the available tools. Coupling constraints are about the location, time and duration the individual has to join other individuals, tools or materials. Finally, the authority constraints are the limitations on a domain in time and space given by an individual or group. No evidence was found in the literature of the application or usefulness of these constraints to flow map design. The theory by Hägerstrand has been used for many interactive visualization techniques (Pred, 1977).

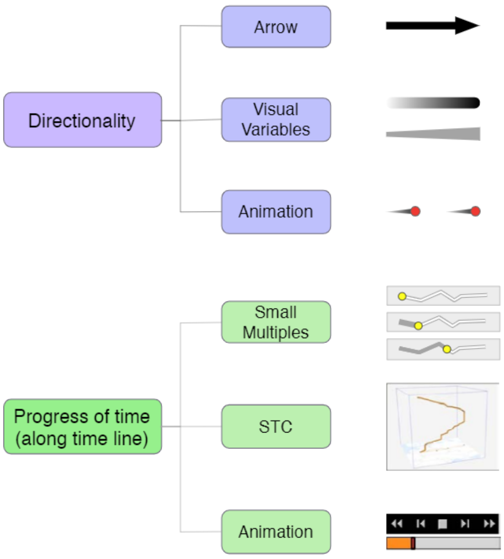
The geographic scale should be chosen in line with the research goals. "*The scale of spatial analysis is reflected in the size of the units in which phenomena are measured and the size of the units in which the measurements are aggregated*" (Andrienko et al., 2010). Visualizing in a different scale could affect the usability results, because the user could interpret the sense of distance differently. As described in Tobler's first law of geography '*everything is related to everything else, but near things are more related than distant things.*' (Tobler, 1970, p. 236). This type of spatial dependence can enable interpolation or extrapolation and spatial or temporal interference. Just like the geographical scale, the temporal scale can distort the outcomes of the map interpretation. Figure 2.5 illustrates three options to display time in a legend. According to Andrienko et.al., GIS scientist have been weak in dealing with temporal data analysis. The concept of time has usually been considered as a linear progression, although time has an semantic structure (2010). Multiple approaches of time exist, such as cyclic time instead of linear time, time intervals instead of time points and time with multiple perspectives instead of ordered time. Using animated maps would create for more options to display dynamic spatio-temporal data.

Figure 2.5: Temporal legends (Harrower & Fabrikant, 2008)



As flow maps are representing movement between at least two locations, the directionality of the flow line can be visualized in various methods. Jenny et.al. (2018) stated that using arrows to represent directionality is the most effective visualization method, although these findings are based on the visualization of 2D static flow maps. Dynamically moving particles in an animated flow map could indicate the directionality as well (figure 2.6).

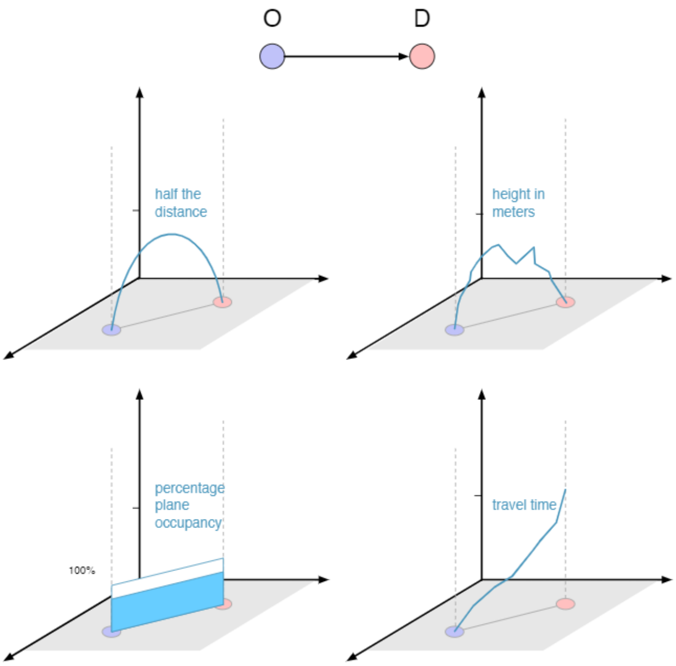
Figure 2.6: Visualization options for the directionality and progress of time (Gu et al., 2017)



Traditionally, motion has been displayed with frames representing individual time fragments, but today's technologies are enabling more dynamic cartographic visualisations (Johnson & Nelson, 1998). Animation improves the ability to study and remember trend patterns (Harrower & Fabrikant, 2008). The user should have control over the animated sequencing and colour coding quantities to make the attributes more distinguishable (Johnson & Nelson, 1998). Animated maps are an interactive data type because the data of the different time segments can be compared and explored by the user. Animation is an example of a technique for adding more dynamism to a map.

When visualizing flow maps in a 3D environment, the third axis is used to represent time in most spatio-temporal data visualizations, but it can be used to represent the quantitative attribute information of a flow line as well (Seipel, 2013). In figure 2.7, four examples of a 3D visualization of flow trajectories is shown. The third dimension offers more design options for flow maps while at the same time it may lead to an increase of the cognition load (Gu et al. 2017). This way, maps can be created with flows appearing as three dimensional lines above a map layer or globe. However, the usage of 3D for geovisualisation can be challenging because not all cartographic theories are developed to be applicable to this visualization type. Specific challenges in 3D visualization exist, such as occlusion (objects are hidden behind other objects), visual clutter and the absence of map scales (Semmo, Trapp, Jobst, & Döllner, 2015). As mentioned in the introduction, a 3D visualization is a dynamic representation of an object in an immersive environment with three axis (x, y, z). The immersive element allows the user to navigate around the visualized object in a virtual space (Read et al., 2010).

Figure 2.7: Examples of flow line symbology of flow lines in 2.5D. (Gu et al., 2017)



2.3.3 Available solutions for minimizing visual clutter

Although no solutions for visual clutter are existing today, there are some options that could reduce it. As an example, curved lines could minimize visual clutter because this could limit the amount of lines passing through areas and/or nodes (figure 3). The design principles listed in table 2.1 are reducing visual clutter for 2D flow maps (Jenny et al., 2018).

Table 2.1: Overview of all visualization techniques for reducing visual clutter in 2D flow maps. Jenny et al, 2018)

Number	Visualization technique
1	minimizing overlaps
2	symmetric flows are preferred to asymmetric flows
3	longer flows are curved more than shorter/peripheral flows
4	acute angles between crossing flows are avoided
5	sharp bends in flow lines are avoided
6	flows do not pass under unconnected nodes
7	flows are radially distributed around nodes
8	flow direction is indicated with arrowheads
9	flow line volume is scaled with represented quantity

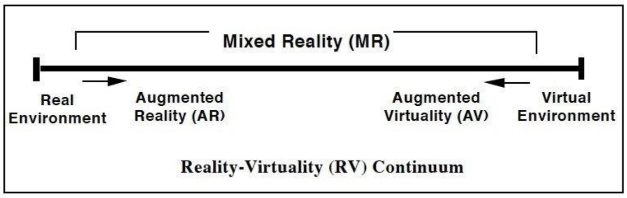
2.4 The development of modern techniques in cartography

As of today, most maps have been created as a static 2D visualization intended for print or web, and increasingly in 2.5D visualization as well. Cartographic rules have traditionally been intended for 2D visualizations, but have been adapted to 2.5D visualizations as well. New techniques such as augmented- and virtual reality allow the user to be emerged in a dynamic 2.5D environment, since you can move freely around the virtual objects. The freedom to navigate yourself in all directions around a model would empower the user to use a more interactive navigation, compared to the traditional PC controls (Seipel 2013).

The currently available techniques for displaying virtual objects in a 2.5D environment are virtual reality, augmented reality, a 3D screen or stereoscopy. VR and AR can create an immersive user experience, while a 3D screen only allows the user to perceive depth without the need to use 3D glasses. The immersive capability creates a high level of map interaction, which is beneficial for an effective visualization (Carbonell Carrera & Bermejo Asensio, 2017). This way, the user can move around the virtual environment to explore the 2.5D visualization from any desired perspective.

Mixed reality (MR) is a spectrum of possibilities for displaying and perceiving information, as explained by the reality-virtuality continuum (figure 2.8) (Milgram & Kishimo, 1994). Objects are either real or virtual; “*Real objects are any objects that have an actual objective existence, whereas virtual objects exist in essence or effect, but not formally or actually*” (Milgram and Kishimo 1994). Although a real object can be observed directly, a virtual object must be simulated. For cartographical purposes, augmented reality is mainly covered by physical geography research papers. However, instead of interpreting the landscape relief or underground structures, AR can be applied to flows maps as well as a means of MR to explore and analyse demographic, cultural or economic movements (Carbonell Carrera & Bermejo Asensio, 2017). An example of a usability study about printed versus AR maps, resulted in a 93% score of the participants who preferred the AR system above the printed maps (Henrique, Pereira, Stock, Stamato Delazari, & Centeno, 2017). Virtual maps have more visualization advantages and opportunities compared to static paper maps.

Figure 2.8: Miligram’s reality-virtuality continuum (Milgram & Kishimo, 1994)

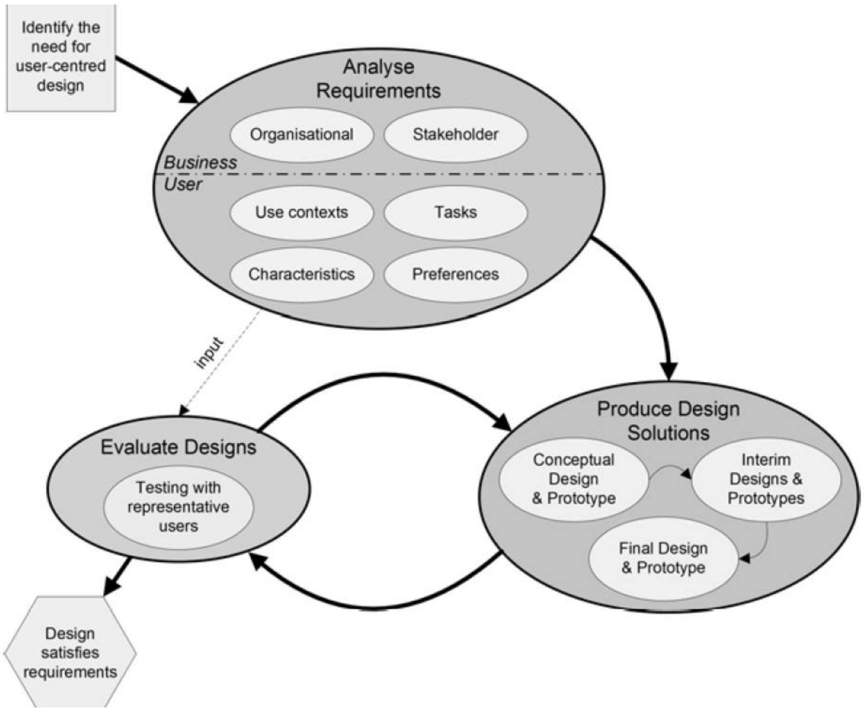


2.5 Usability study

As mentioned earlier, the main goal of today's cartography is to create maps as a medium for communication of information, as well as to make exploratory data analysis more widely accessible (Andrienko, et.al. 2010). "*The choice of a graphic characteristic is crucial for a correct transfer of contextual meaning to the map user, and that symbols designed according to aesthetic rules will hold the user's attention to the map for a longer period of time*" (Haeberling, 1999). This quote demonstrates the importance of cartographic principles for a good map usability. The involvement of users during the design phase is an important design aspect, since critical feedback can be received in an iterative visualization design process to improve the final product (Semmo et al., 2015).

Jenny explained how efficiency, effectiveness and satisfaction are three main elements that can be used to measure the map usability quality (et.al., 2018). The efficiency relates to the readability speed as the necessary time to retrieve the information from the map. The effectiveness relates to the correct transfer of the information, which is measured by checking whether the answers in the usability test are correct or not. The satisfaction relates to the experience of pleasantness when reading the map (Roth 2017). These three usability characteristics are measured individually, so these scores can be analyses in comparison with each other after the usability test has finished.

Figure 2.9: user-centred design process (based on Jokela et al. 2003; and Williams and Lafreniere, 2005. (van Elzakker, 2004)



2.5.1 Usability and readability

The definition used for visual clutter is: "*...a limitation of the map readability caused by intersection or clustering of multiple flows*" (Jenny, et.al. 2018). Assumptions on visual clutter in the usability test are made by testing the usability and readability based on the participant's behaviour, their think aloud commentary and the usability questionnaire results. For clarification, the following definitions are used for the terms usability and readability. It should be noted that both terms are different from visual clutter.

For ' usability', the following definition by the International Organization for Standardization (ISO) on the *Ergonomics of human-systems usability interaction* is used: "*The extent to which a system, product or service can be used by specified users to achieve specific goals with effectiveness, efficiency and satisfaction in a specified context of use*" (International Organization for Standardization, 2018). Usability is a measurement for defining the user experience of a product. The three main outcomes of a usability test are; it should be easy to become familiar, it should be easy to achieve the objective and it should be easy to recall the user interface for subsequent visits. The three terms effectiveness, efficiency and satisfaction are used as a measurement for the map usability.

Map readability is determined by a composite of measures. It is influenced by the level of visual complexity, which includes the amount of information, the object complexity (proper level of generalization), the graphical resolution (suitable symbol style) and the spatial distribution of information (Harrie et al., 2015). A good visualization should be simple, in which these four measures effectively complement each other for avoiding visual distractions as much as possible. Confusion will hinder the communication of information from the map to the user (Johnson & Nelson, 1998). A reduction of visual clutter is not equal to an improved map readability. However, a better readability could cause a reduction of visual clutter. This correlation is tested with qualitative research method, which is explained in the next section.

3. Methodology - design

3.1 Introduction chapter

The usability test is the data gathering component of this research and is divided into two separate chapters. In this chapter, the research method, usability test workflow, map design, questionnaire design, focus group preparation and the final script of the usability test is described. The chapter ends with the final preparations and challenges for the performance of the usability test, which is explained next in chapter four.

3.2 Research method

3.2.1 Qualitative research method

A qualitative research methodology is used for this research. Instead of referring to counts of measures only, which can be statistically analysed as means of quantitative research, a different approach is applied. Qualitative research is based on the observation of non-numerical data, for exploring phenomena and describing individual experiences (Farr, 2008). This way, more flexibility is available for the information collection, specifically for interviews or discussions where respondents can add additional information instead of answering identical questions in the same order. This added information can be meaningful and explanatory for the research, which is beneficial for the analysis.

The qualitative approach of an interview section at the end of the usability test and a think aloud method for providing comments during the test is combined with a structured questionnaire with closed questions. Because of the scope limitations of this research, a sample group size ranging between 16 and 24 is expected, and therefore no quantitative research measures can be applied to the results.

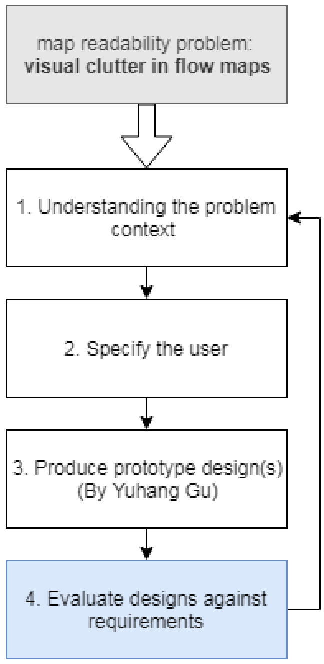
A think aloud method requires all participants to speak out their reasoning while looking at the maps and answering the related questions. By recording their voice as a backup, and making notes during the sessions, additional information can be collected that otherwise would have not been retrieved with a closed questionnaire method (Olmsted-Hawala, Murphy, Hawala, & Ashenfelter, 2010). With this method, the participants are not distracted while sharing their reasoning, which is valuable information for the usability investigation.

3.2.2 User centred design

A usability analysis enables the early involvement of the user during the product development; end users or experts can provide feedback during the initial production process (Bevan, Carter, & Harker, 2015). This is an effective and efficient method for creating a product or service that effectively suits the needs of its users. User centred design (UCD) is a philosophy based on the needs and interests of the user, with an emphasis on making products usable and understandable (Nielsen & Landauer, 1993). The four key phases of the UCD approach are illustrated in figure 3.1. Likewise, the results from this research can be used for further improvement of the prototype maps.

The problem context is elaborated with a literature study, the user group is specified in relation to the produced prototype. After completion of UCD component 1 and 2, the list of requirements for the prototype can be set up. In collaboration with Yuhang Gu will be discussed if adjustments need to be made before the model is ready for the user study. The fourth UCD phase, is the design evaluation phase. The results of the usability study can be used for the further refinement of the prototype maps and for follow up research with a larger sample size for investigating the usability effects of map stimuli in flow maps in relation to visual clutter.

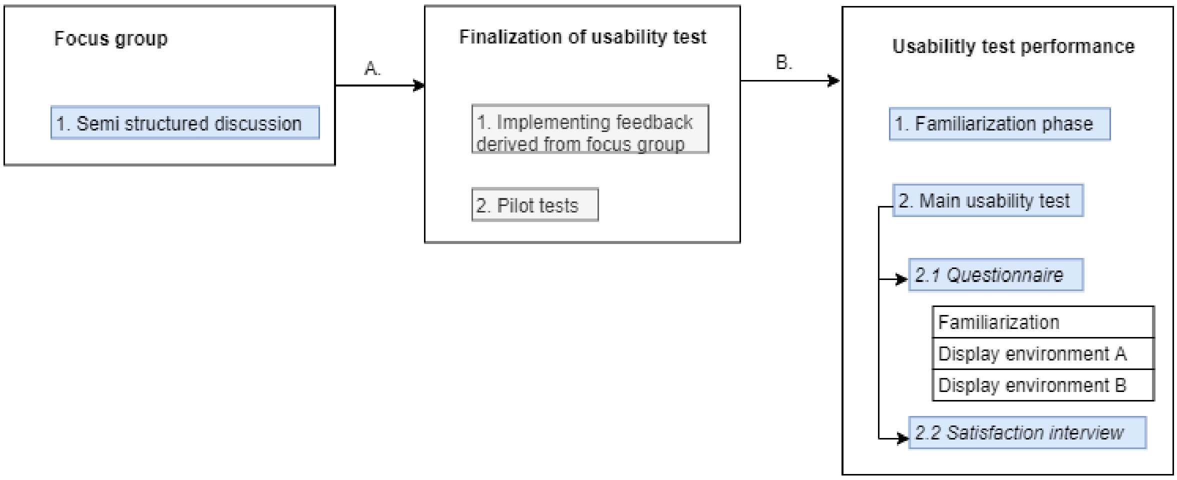
Figure 3.1: The four main phases of the user centred design approach (based on ISO 1340: 1999).



3.2.3 Usability test workflow

The usability study consist of two main components; a focus group discussion with domain experts in cartography, data visualization or user experience design, and a usability tests performed by predominantly students in the faculty of geosciences (Boeije, 't Hart, & Hox J., 2009). Because of the scope of the thesis, the investigator was not able to collect data from more than 30 participants for the usability test. Therefore, only descriptive statistics can be applied to the results for the data analysis.

Figure 3.2: Scheme of the usability test workflow used for this research



The focus group is a qualitative research method in which the researcher is acting as the interviewer and chairman of the discussion. The focus group is expected to result in detailed information about the expert's opinions on the proposed method and prototype for the usability test (McDonagh-Philp & Bruseberg, 2000). A focus group approach is more suitable than individual interviews since these groups often generate rich contextual data within a short amount of time than is needed for individual interviews. A range of opinions and issues can be addressed quickly and the discussion will provide clarification and justification since the respondents can comment on each other's thoughts directly. A focus group will enable less in depth information or personal experiences, and group dynamics can be more difficult to analyse, but this level of personal detail is unnecessary for this research. Between two to five respondents will participate, who are selected on their expertise with cartography, map user experience and/or interactive map design. The audio is recorded for the transcription and analysis. All received feedback from the focus group can be used to finalize the preparation of the usability test (Adams & Cox, 2008).

The usability test is divided into two stages. In stage one, participants are being familiarized with the soft- and hardware in order to start with answering the 48 tasks on the main test by retrieving the information from related maps. The familiarization phase is designed for reducing map reading problems related to the lack of knowledge of the soft-/hardware. For instance, the participant could have difficulty with navigating the MR maps on the iPad, since multiple navigation methods are required and this might be the first time the participant is using this technique. Likewise, difficulties could arise with understanding the map features, such as the legend, flow lines, nodes, arc height, line volume and the flow line colour. The cartographic styling should be optimized for an end user public. Although the population for this test consist of students with experience in map reading, they should be treated as end users by making the maps as accessible as possible. This main test is designed to investigate the flow maps on its effectiveness and efficiency, by analysing the answer quality and the elapsed answer time. By comparing this information with the map stimuli in the corresponding tasks, along with the participant behaviour retrieved from the think aloud and screen recording method, a qualitative analysis is performed on the relationship between the effectiveness-, efficiency and satisfaction results with the four map stimuli.

A quantitative method is not suitable for this usability test because closed questions enable less room to investigate score values for the effectiveness (the answer quality) and efficiency (the speed in which the correct answers are given) (International Organization for Standardization, 2018). Measuring satisfaction is more complicated, since it is based on subjective opinions which could range more widely across the sample group. Although questions on the participant's satisfaction on the maps can be asked in closed questions, the results will be compared to the information derived from the observations.

The researcher is attending all usability tests, and will be observing the participant's behaviour by taking notes, recording the participant's voice for transcribing the think aloud process afterwards and recording the screen of the device used (PC screen and iPad). The diverse set of information collected in parallel is analysed afterwards for a more valuable set of results. The investigator will not interfere with the respondent during the main test, since it could influence the participant's thinking process and therefore their answering behaviour, which could lead to invalid results (Olmsted-Hawala et al., 2010). If the participant has questions, these can be answered by the investigator up and until the end of the familiarization phase.

The complete usability test setup, the recording material and an explanation of the map navigation controls are communicated with the participant by a printed manual sheet at the beginning (appendix A2). By using a written manual for all participants, all participants are presented with the exact same explanation information. The combination of the written instructions and the limited interaction possibilities with the investigator are minimizing the evaluator effect; a substantial difference in what multiple individual evaluators are observing for the same session (Jacobsen, Hertzum, & John, 1998).

The two main topics to be covered by the test questions are based on the four map stimuli; two flow line symbologies and the two display environments. For visualizing quantitative attribute information, a variation in either the arc height or the line volume is applied. Both a 3D map on a screen as well as a MR 3D map on an iPad are used as the display environments.

3.2.4 User characteristics

The user groups for both research components are selected based on their skills and knowledge. For the usability test, students of the faculty of geosciences at Utrecht University are asked to participate because of their experience with map reading and the demographic topic that is addressed in the prototype data: migration between municipalities within the same province in the Netherlands. Map reading experience is therefore highly preferred for the test since the user are required to understand the geographic context of the presented data (Harrie et al., 2015). . However, the test is designed to be suitable for people without a geographical background as well, but there is chosen to limit the participant population to students within the field of geosciences because this research is performed on a small scale.

3.2.5 Research collaboration

Yuhang Gu is a PhD candidate at ITC faculty of geo information science and earth observation. His expertise of building 3D flow maps in a web browser enabled the development of the usability test. This way, deliverables of both this MSc. and the PhD research are complementing one another. See table 3.1 for a list of all his contributions to this thesis research.

Table 3.1: An overview of all contributions to this thesis research by Yuhang Gu

contribution task	description
developing the map prototypes	The prototype was used during the midterm presentation, with a dataset of the country of Iceland as an example.
brainstorming and discussions	During the design, performance and analysis of the usability test, Yuhang was my first contact person for questions and discussions.
developing the final maps and online questionnaire	Based on the requirements of the usability test; questions, answer options, data sources, visualization characteristics and additional text, Yuhang built the 8 maps within an online web interface for viewing the maps, along with an additional online questionnaire. Yuhang executed this technical development, whereas the structure of the questionnaire, content of the maps and the questions were designed by the author.
first analysis	Since only Yuhang had access to the data of the usability test results, Yuhang created the charts for the analysis of the usability test (efficiency and effectiveness).

3.3 Map design

3.3.1 General

3.3.1.1 Data preparation

The data source used to represent the flow lines is called 'Migrated people within and between municipalities' (*Verhuisde personen binnen gemeente, tussen gemeenten*), derived from the Central Bureau of Statistics (CBS, 2017). Likewise, data from 2017 is used for the base map layer with the municipality boundaries. The flow lines are divided amongst four classes based on a mean classification method. This way, four different line volumes are applied to the flow lines to represent the same range of migrating people as for the variations in the maps symbolised with arc height.

For the familiarization phase, municipalities in only the province of Utrecht are used, since this phase is designed to train the participant with the soft- and hardware of the map product, to reduce any potential error or delays in the main test caused by unfamiliarity with the techniques. However, when the same geographical area is used for the usability test, the participants are likely to answer the upcoming tasks more accurately because the participant is expected to become familiarized with not only the techniques, but with the data source as well. The province of Utrecht is used for the familiarization, because all participants are students from Utrecht University. Therefore it can be assumed that the participants generally have a better understanding of this geographical area compared to the rest of the Netherlands. For time constraints, only two maps, one for each environment, are used for the familiarization.

For the main test, two other geographies are used for the maps: the municipalities in the province of Friesland and Groningen. Since the relative difference of the migration sums over the years is low, and for the simplification of the usability test, all data is used from 2017 only, the most recent year available. an alternation of two maps with different data sources will reduce the speed in which the participant is expected to become familiarized for avoiding invalid results as discussed above. The selection criteria for the two provinces are that the geographical shape and the number of municipalities should be as similar as possible. Friesland and Groningen are chosen since both provinces have a similar amount of municipalities (24; 23 municipalities). All provinces have a local scale level in relation to the world, therefore minimal distortions effects of the circular shape of the Earth's surface occur.

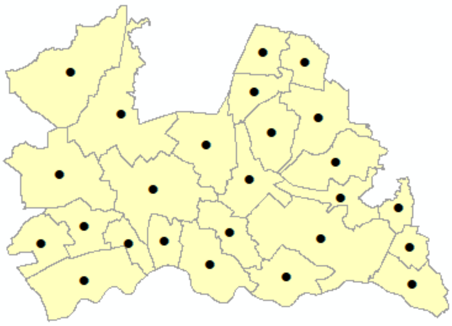
3.3.1.2 Map features

The design requirements are equal for all maps throughout the usability test. Because the research goal is testing the usability based on four map stimuli, no other map features are included in the flow maps for preventing the participants behaviour to be influenced by other features. The following map features are implemented: flow lines, flow line bar legend including the corresponding class numbers, municipality label numbers and the buttons for switching between the maps. It is chosen to exclude a scale bar because the sense of geographical size is less relevant for the purpose of this study. For the same reason, the legend information on the flow line directionality is explained in the explanation box in the online questionnaire (red lines are going out, blue lines are going in to the selected municipality).

For the base map, the municipalities are visualized as polygons. Its centroids are the origin- and destination locations of the flow lines (figure 3.2). The municipality boundaries used are from 2017, which is the same year as the CBS migration data used for the flow lines. One province is shown per map, including all corresponding municipalities per province (number of municipalities per province; Utrecht: 26, Friesland: 24 and Groningen: 23).

All flow lines are only curved on the vertical axis and are either styled by colour to indicate the direction of the flow lines. Blue lines represent flows that go in to the focus municipality, red lines represent the flows going out. A legend for this direction is not included in the map viewer for reducing the amount of map features. Instead, it is explained during the online questionnaire and for all questions, the directionality is explained by stating the set of origin- and destination municipalities like this: (3 → 10), representing a flow from municipality 3 to municipality 10.

Figure 3.2: Example of map with centroids for each municipality within a province.
The flow lines between municipalities will originate and end at these point features.



3.3.2 Display environments

The ability to navigate a 3D map by moving the tablet computer around a table is an example of an interactive map because it enables access to alternative layouts of information on demand (Wood, Dykes, & Slingsby, 2010). However, 3D maps are creating both opportunities and challenges. Therefore, alternative solutions are needed for flow maps. In particular, the combination of 3D with augmented reality could result in unforeseen usability difficulties as well, since this techniques has not been researched for flow maps before. The goal of this thesis research is to investigate the potential added benefit of an enhanced usability when visualizing flow maps in 3D with MR, compared to a PC screen.

Digital maps could have the ability to interact (i.e. pop-up attribute information) and navigate (panning/zooming). Although dynamic maps created in 3D can be developed with more tools and visualization techniques than for static 2D maps, it can not be assumed that 3D maps are instantly providing a map with less visual clutter and a better readability. For validity purposes, the 2D display environment is not integrated in the usability test, since a simplified number of map variables contributes to a more valid experiment. Therefore, the two rows coloured in red in table 3.2 about the 2D map characteristics are indicating that this option is not incorporated in this research. The longer the duration of the test, the higher the chance that participants will become less focused. Additionally, using less map variables provides more opportunity for a reliable analyzation of the results.

Since the majority of the available academic knowledge on flow map design is based on two dimensional maps, most solutions for visual clutter in 2D maps can not be directly applied to 3D maps. For example, sharp angles should be avoided and symmetrically shaped lines are better than asymmetrical lines for 2D flow maps (Jenny et.al., 2016). Jenny stated that curved flow lines is an example of a representation which is proven to be more immune for the visual clutter problem in a 2D environment. However, curved lines are difficult to visualise effectively in 3D since lines can be observed from all angles and are curved to the z-axis already. According to Jenny, the purpose of curved lines is to reduce clutter created by overlapping lines and nodes. While 2D maps are observed from the same view (from above), the user navigates around a 3D flow map from any angle (Jenny et al., 2018).

Another flow line symbology that has proven to reduce visual clutter in 2D flow maps is the usage of arrowheads instead of tapered line volumes for indicating the flow direction (Jenny et.al., 2016). Arrowheads will not be applied to the maps for the usability test since its effectiveness for reducing visual clutter is proven and the usage of a different display environment is unlikely to have a different effect (Jenny et al., 2018).

Figure 3.3: Flow map concept in a PC environment. The map is perceived on the screen.

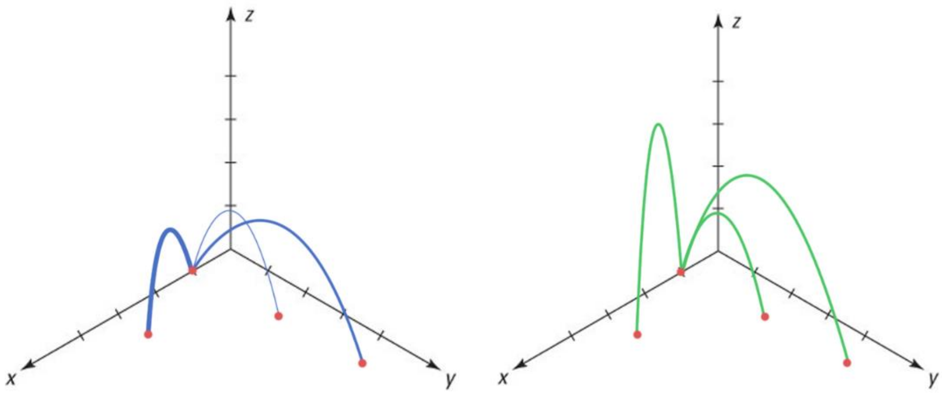
Figure 3.4: Flow map concept in a MR environment. The map is perceived on an iPad, aimed at the target on a flat surface such as a table.

3.3.3 Flow line symbology

Similarly to the display methods, only two flow line symbologies are integrated in the test: line volume and arc height. These variables are chosen because both can be applied to represent quantitative attribute information for individual flow lines. A uniform colouring is applied as the symbology for indicating the difference in directionality. Blue lines represent the direction going in, red for the direction going out (table 3.5 and 3.6). These colours are randomly chosen selected.

Since all maps are displayed in 3D, all flows are visualized as vertically arched lines with an x-, y- and z-coordinate. However, the arcs have equal heights for the maps covering line volume, whereas the arc height is varied in the second flow line symbology but with equal line volumes (table 3.5). The map stimuli line volume is chosen for validating whether the results on the effectiveness of a line volume styling by Jenny can be applied to multidimensional MR maps as well, since it was tested on 2D maps only (Jenny et al., 2018). The cartographic stimuli height is chosen because it is exclusively compatible to a 3D environment, since the z-axis is required to show the height differences on the map. Therefore, research is needed for a better insight in the potential added benefit of using height differences as a three dimensional styling option in flow maps in a 3D (MR) display environment.

Figure 3.5: The two flow line symbologies used in this research.
line volume variation (left, in blue) and arc height variation (right, in green)



Both flow line symbologies are symbolizing absolute quantitative attribute information. For the first symbology, a thicker line volume is representing a higher (absolute) attribute value and a higher arc height is used to indicate a higher attribute value (figure 3.5). For objectivity purposes, the maps are as similar as possible between the environments. This means that certain functionalities are left out, because these are not (yet) available for MR or because of the time limitation of this research. An example is the interaction with map features, which could allow the user to click on municipalities or lines to select individual lines by applying a glow effect to the line, retrieve more information about the quantitative flow size by a pop up label/window or to switch between the 'focus' municipality themselves. This functionality is readily available on a PC environment, but it would be technically challenging to apply this functionality to MR. An suggestion would be to have an target fixed in the centre of the iPad screen, and the user needs to aim the camera so the desired flow is pointed at by the target, and then click on a button to 'retrieve more information or highlight it'. This could be an interesting capability, because it further increases the capabilities of flow maps in this map environment which could increase the usability, but this is a topic outside of the scope of this research.

3.3.4 Map navigation

Other than the mentioned visual elements are navigation controls regarded as features as well. For interacting with the map, the controls are different depending on the display environment used. See table 3.2 for an overview of the navigation controls.

Table 3.2: Navigation instructions for the maps in both display environments

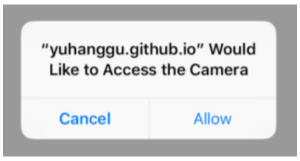
	display environment	
tasks	PC monitor	MR iPad
open the map	1. click on the map symbol	1. tap the map symbol 2. allow the following notification: 
return to menu	click on return page (Chrome) 	click on return page (Safari) 
rotate orientation	mouse left click and drag	move iPad device, with the camera aimed at the target placed on the table
pan	right click and drag	
zoom	turn the scroll wheel	

Table 3.3: The usability characteristics per flow line symbology applied in the usability test

usability characteristics per flow line symbology	
line volume	arc height
3D maps	
uniform arc height	varied arc height
varied line volume	uniform line volume
2D maps	
Line volume variation is a flow line symbology that has been applied to 2D maps already. When additional attribute data is added, the lines can have multiple colours, otherwise a uniform colouring is used.	It is impossible to display arc height in 2D flow maps, because the map is viewed from above.

Table 3.4 The usability characteristics per display environment applied in the usability test

usability characteristics per display environment	
PC	MR
viewing from any perspective	viewing from any perspective
interaction with data is possible, but not included	no data interaction possible
navigation by panning/zooming	navigation by moving the device

3.4 Questionnaire design

3.4.1 Stimuli sequencing

The usability test questionnaire is based on the classification size and the geographical location or spread of the flows in its corresponding map for testing the information readability and the map usability. An equally balanced division of both PC and MR maps used is necessary for a valid analysis between the three mapping environments. In total, two maps are used for the familiarization phase and sixteen maps for the usability test. The ordering of the maps is based on the four characteristics. However, only the two map stimuli are used for the analysis of the results See table 3.5 for an overview of the map characteristics.

Table 3.5: Overview of the flow map characteristics used in this research

	flow map characteristics	
Map features used for the usability analysis (= map stimuli)	**display environments** • PC screen • MR on iPad	**flow line symbology** • arc height • line volume
Map features used for the presentation of a variation of data	**dataset** • Friesland • Groningen	**flow line direction** • in (blue) • out (red)

When using a single questionnaire version for all participants, a comparison between different map features is more complex to analyse. This is because the user might become familiarized with the data or a specific map display option, which might lead to biased answers. For instance, asking about information on a map from the same data source could be easier to understand when presented again but with differently applied features. Likewise, when the participant is provided with maps with *stimuli X* first, it might have an effect on the speed and quality of the answers at a later stage with the other map stimuli. It is essential to consider this learning bias effects, so it can be prevented in order to achieve results with a higher research validity. Table 3.6 illustrated all maps used in the main usability test, divided by display environment, data source and flow line symbologies.

By dividing the total number of maps in two main groups based on the display environments, the participant only has to switch between devices once during the main test. The minimization of switching between the PC and the iPad makes the test structure more practical because it reduces the total test duration time. The subgrouping is based on the dataset (Friesland or Groningen), in order to test the potential effect of the difference of geography. Both the flow line symbologies and the directionality will be presented alternately.

Therefore, sixteen questionnaire versions with multiple map stimuli sequencing applied are used for a more effective comparison of the map stimuli and better validity of the results. This task ordering design is inspired by the a matrix model, the 'Latin Square' theory used in statistics (Andrienko & Gennady, 2006). A stated by Macneish (1922): "*A Latin Square of index n, k gives a schedule for a contest between k teams of n members each, where each member is to meet each member of the other teams precisely once, and each member is to participate but once at each field*". The versions are equally divided amongst all participants. See table 3.7 for the overview of the questionnaire ordering of all maps.

Table 3.6: Overview of all maps used for the familiarization and main usability test, divided in main- and subgroups

phase	group	data	sub group	map code	city name	city number	representation	direction
familiarization	-	Utrecht PC	H	TU1	Zeist	19	height	*in*
familiarization	-	Utrecht AR	W	TU2	IJsselstein	18	volume	out
main test	A	Friesland	H	1.AF1	Leeuwarden	9	height	*out*
main test	A	Friesland	H	2.AF2	Smallingerland	14	height	*in*
main test	A	Friesland	W	3.AF3	Tietjerksteradeel	19	volume	out
main test	A	Friesland	W	4.AF4	Súdwest Fryslân	22	volume	in
main test	A	Groningen	H	5.AG1	Slochteren	14	height	*out*
main test	A	Groningen	H	6.AG2	Hoogezand-Sappemeer	9	height	*in*
main test	A	Groningen	W	7.AG3	Grootegast	1	volume	out
main test	A	Groningen	W	8.AG4	Haren	13	volume	in
main test	B	Friesland	H	9.BF1	Franekeradeel	5	height	*out*
main test	B	Friesland	H	10.BF2	Leeuwarden	9	height	*in*
main test	B	Friesland	W	11.BF3	Ooststellingwerf	7	volume	out
main test	B	Friesland	W	12.BF4	De Fryske Marren	24	volume	in
main test	B	Groningen	H	13.BG1	Veendam	15	height	*out*
main test	B	Groningen	H	14.BG2	Leek	10	height	*in*
main test	B	Groningen	W	15.BG3	Delfzijl	23	volume	out
main test	B	Groningen	W	16.BG4	Winsum	17	volume	in

Table 3.7 Sequencing of the maps per user ID. (columns in blue are indicating the PC environment for more clarification)

User ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Map	TU1	TU1	TU1	TU1	TU1	TU1	TU1	TU1	TU1	TU1	TU1	TU1	TU1	TU1	TU1	TU1
	TU2	TU2	TU2	TU2	TU2	TU2	TU2	TU2	TU2	TU2	TU2	TU2	TU2	TU2	TU2	TU2
	PC	**MR**	**PC**	**MR**	**PC**	**MR**	**PC**	**MR**	**PC**	**MR**	**PC**	**MR**	**PC**	**MR**	**PC**	**MR**
	AF1	AF1	AF3	AF3	AF1	AF1	AF3	AF3	AG1	AG1	AG1	AG1	AG3	AG3	AG3	AG3
	AF2	AF2	AF4	AF4	AF2	AF2	AF4	AF4	AG2	AG2	AG2	AG2	AG4	AG4	AG4	AG4
	AF3	AF3	AF1	AF1	AF3	AF3	AF1	AF1	AG3	AG3	AG3	AG3	AG1	AG1	AG1	AG1
	AF4	AF4	AF2	AF2	AF4	AF4	AF2	AF2	AG4	AG4	AG4	AG4	AG2	AG2	AG2	AG2
	AG1	AG1	AG1	AG1	AG3	AG3	AG3	AG3	AF1	AF1	AF3	AF3	AF1	AF1	AF3	AF3
	AG2	AG2	AG2	AG2	AG4	AG4	AG4	AG4	AF2	AF2	AF4	AF4	AF2	AF2	AF4	AF4
	AG3	AG3	AG3	AG3	AG1	AG1	AG1	AG1	AF3	AF3	AF1	AF1	AF3	AF3	AF1	AF1
	AG4	AG4	AG4	AG4	AG2	AG2	AG2	AG2	AF4	AF4	AF2	AF2	AF4	AF4	AF2	AF2
	MR	**PC**	**MR**	**PC**	**MR**	**PC**	**MR**	**PC**	**MR**	**PC**	**MR**	**PC**	**MR**	**PC**	**MR**	**PC**
	BF1	BF1	BF3	BF3	BF1	BF1	BF3	BF3	BG1	BG1	BG1	BG1	BG3	BG3	BG3	BG3
	BF2	BF2	BF4	BF4	BF2	BF2	BF4	BF4	BG2	BG2	BG2	BG2	BG4	BG4	BG4	BG4
	BF3	BF3	BF1	BF1	BF3	BF3	BF1	BF1	BG3	BG3	BG3	BG3	BG1	BG1	BG1	BG1
	BF4	BF4	BF2	BF2	BF4	BF4	BF2	BF2	BG4	BG4	BG4	BG4	BG2	BG2	BG2	BG2
	BG1	BG1	BG1	BG1	BG3	BG3	BG3	BG3	BF1	BF1	BF3	BF3	BF1	BF1	BF3	BF3
	BG2	BG2	BG2	BG2	BG4	BG4	BG4	BG4	BF2	BF2	BF4	BF4	BF2	BF2	BF4	BF4
	BG3	BG3	BG3	BG3	BG1	BG1	BG1	BG1	BF3	BF3	BF1	BF1	BF3	BF3	BF1	BF1
	BG4	BG4	BG4	BG4	BG2	BG2	BG2	BG2	BF4	BF4	BF2	BF2	BF4	BF4	BF2	BF2

3.4.2 Questions

3.4.2.1 Difficulty levels

All tasks can be distinguished between two difficulty levels, which are based on the question complexity. For the elementary level, only one data instance is used for the task; one set of two municipalities with flows in between, one for each direction. The other level is called synoptic, which is a combination of all other difficulty levels (intermediate and overall) (Andrienko & Gennady, 2006). Synoptic tasks will require the participant to consider the dataset in its entirety; all flows between all municipalities for each direction. Two different subcategories exist for the synoptic level: *connective* and *descriptive.* Whereas descriptive tasks are only explaining the nature of the data items, the connective tasks requires the user to further analyse the data on the relationships between flows and/or nodes (Andrienko & Gennady, 2006).

Table 3.8: OD matrix with the level two connections marked with the red row and column

Finish → -------------- Start ↓	municipality A	municipality B	municipality C
municipality A	X		
municipality B		X	
municipality C			X

As explained above, the tasks are divided by two levels; elementary (1) and synoptic (2). See the full list of questions per map in appendix A1. Instead of multiple choice answer options, the participant is asked to fill out the answers themselves in plain text. This way, the participant can only find the answer in the map, instead of comparing between the answer options for a multiple choice question type. For each of the sixteen maps in the usability test, three questions are asked. One question has the elementary level, the other two questions are synoptic. An equal division is made for the direction of the flows (8 maps with flows going in, 8 maps with flows going out). As an example, the questions corresponding to map TU1, the first map in the familiarization phase, are shown below (figure 3.6).

Figure 3.7: Usability test, map TU1

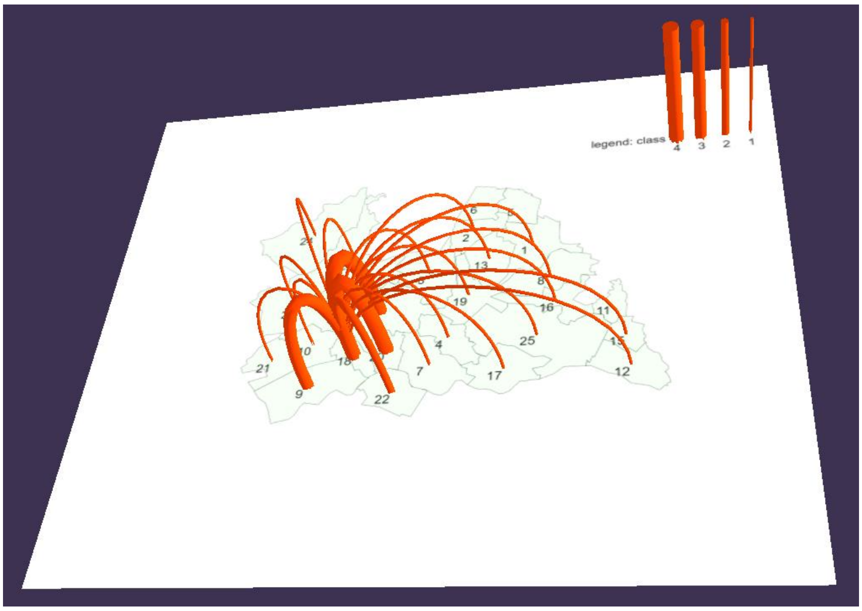


Figure 3.8: Usability test questionnaire, map TU1

Compare flows from municipality 18 to municipality 20 (18->20) and municipality 22 (18->22), which is the larger?

18->20

18->22 equal

To which class does the flow from municipality 18 to municipality 26 (18->26) belong?

1
2
3
4

List two destination municipalities with the largest flows departing from municipality 18.

< Write down number of municipalities (such as: 1, 2, 5). >

3.4.2.2 Scoring method

For measuring usability, three components are analysed; efficiency, effectiveness and satisfaction. The extent to which the participant has answered the questions correctly relates to effectiveness, the answer time to efficiency. Specifically formulated questions on the participants preference of the presented map types is relating to satisfaction. The questions are targeted to cover each main feature equally during the test: flow line symbology and display environment.

An online questionnaire is used, which the users can fill out on their smartphone. The questionnaire will measure the elapsed answer time for each map. The satisfaction of eight of the sixteen maps is measured by a short Likert scale survey. These eight maps are selected based on a single direction, therefore an equal number of maps in both display environments and flow line symbologies is covered. The investigator will ask the users to explain their answers. The answer time can therefore be analysed in seconds, per map. The effectiveness is based on a score for each question per map. The total score per map ranges between 0 and 3.

- Question 1: question 1 is making choice from the provided options, when participant make the right choice, this question is scored as 1 point.
- Question 2: the same as question 1.
- Question 3: this question require participants find one or two flows. When the correct answer contains 1 flow, the rule is the same as question 1 and question 2. When there are more than 1 flows in correct answer: the participant will scored 0.5 point for each correct flow.

3.5 Focus group

3.5.1 Preparation

A focus group is an informal discussion organised by a moderator who should insure not to direct the discussion, for a more active behaviour of the participant for an increased interaction between the participants (McDonagh-Philp & Bruseberg, 2000). The discussion leader of a focus group should moderate the structure of the session, try to let the participants take the lead of the discussion and to prevent to let dominant speakers have a too large influence in the discussion. The interactive nature of a group discussion enables the participants to listen to others views, which might results in a more nuanced response compared to an individual interview. The group consists of carefully selected people with a shared commonality which is of interest by the researcher (Parker & Tritter, 2007). The focus group interview is expected to generate critical feedback on the usability test methodology including the prototype maps. Depending on the feasibility of the received points of feedback, the usability test methodology is refined one last time. Focus group sessions are especially effective for product development cycles, in which a prototype can be efficiently critiqued by a group of experts before the production process begins. "*The success of a product is not only substantiated by the decision of customers to purchase it, but also by the satisfaction and pleasure gained through its ownership*." (McDonagh-Philp & Bruseberg, 2000).

Instead of taking notes, audio recording is used to enable transcription afterwards (see appendix A5). For ethical and privacy concerns, the participants need to agree with the usage of the audio recording before the start of the focus group session. Additionally, the transcripts of the focus group session will be send to all respondent, in which they individually can agree with the statements made and if not clarify certain elements. This is both out of ethical reasons towards the respondents as for the validity of the results

A semi structured approach is used for the preparation of the session. Beforehand, the participants are informed with the problem context of this research and a short topic list. The iPad Pro device is used to display the MR maps in this session as well. See the topic list below:

Figure 3.9: Focus group topic list

- problem context
- general methodology of the usability test
- prototype maps
- prototype online questionnaire
- additional questions

The participants are experts in cartography, data visualization and/or user experience design. The following two people have participated:

- **Corné van Elzakker** is professor at ITC Enschede. His main research interests are in the fields of use and user issues in geo-information processing and dissemination, including cartography and visualization.
- **Ieva Dobraja** is a PhD candidate in geo-visualization.

3.5.2 Results

The following feedback is implemented in the usability test based on the received feedback from the focus group session (appendix A4):

- The colour of the legend is equal to the flow lines.
- The legend classes are numbered.
- Thinking about how to apply the terminology of the display environments in the thesis and to the participants. Although the terms 'MR' is used for the techniques and the display environment in this thesis, the participants are asked for their experience with AR maps, because the latter term is better known by the sample population. This is why the definition AR is applied to the questionnaire and online maps.
- Present a manual at the start, so all participants are instructed equally.
- Present all maps on the same screen extent, to prevent unwanted usability differences based on the screen size between the devices. Therefore, for PC monitors with a screen bigger than 13", the web browser window (in which the maps and questionnaire are displayed) is fitted to the same dimensions as the iPad.

3.6 Usability test script

This script is used as a chronological summary of the usability test design, which is used as a guide during the performance of the test. The estimated total duration of the usability test is 45 minutes.

3.6.1 Introduction

(duration: 5 minutes)

The participant is welcomed and thanked for their cooperation to this research. The investigator gives a brief explanation of the guidelines and structure of the usability test and asks the participant to read the manual (appendix A2). After the participant is finished reading, he/she is asked for the consent of using the recording materials. Ask if the participant has any questions, if not, he/she can proceed with the familiarization.

3.6.2 Familiarization

(duration: 10 minutes)

This phase is designed to familiarize the participant with the hardware and software used to display the created map content. Since the maps are dynamic (instead of the more widely known, traditional static map type), the user has to read and additionally interact with it in a different way. During the familiarization, users can experiment with the hardware and software, without learning about different dataset used in the main test. Up and until the completion of the familiarization, the user has a the opportunity to ask questions to the investigator. There will be no verbal communication between the investigator and the participant during the main test. Since most of the participants are students at Utrecht University, it could be that they are familiar with this geographical area. However, it should be assumed that not all students live in or near Utrecht and therefore they have a lower familiarity of this area. For reducing the familiarity in the main test, this province is only used for the familiarization phase.

3.6.3 Main test

(duration: 25 minutes)

After the familiarization phase, sixteen maps with three questions per map are covered in the main test. When the participant has continued after map TU2 in the familiarization, a starting screen indicates that the questions will follow when clicking ‘continue’. As soon as the participant continues here, all recording measurements should be turned on by the investigator (audio, video and screen recording). The answer time will automatically be recorded by the online questionnaire.

3.6.4 Satisfaction interview

(duration: 5 minutes)

The second stage of the usability test is designed to answer the third usability component; satisfaction. After all tasks of section 1 are finished, the satisfaction interview is conducted. The interview consists of eight questions which will take up four minutes in time. The investigator is asking which maps the participant preferred to read, by covering all topics of map features which have been used in section 1 of the test as well.

The satisfaction will be based on the map, not on the questions. Therefore, the user's satisfaction is asked for all four different maps (table 3.9). The participant can assign their rating score for each map, in a Likert scale form. The investigator will ask the participant to further explain their satisfaction rating, which will be noted by the investigator. The following eight questions are designed to measure the map satisfaction of a selection of the maps used in the main test . Please provide an honest answer.

Map X has a good readability and no (signs of) visual clutter:

Table 3.9: Likert scale answer options

map number	strongly disagree	disagree	neutral	agree	strongly agree
1					
2					
3					
4					
5					
6					
7					
8					

3.7 Final preparations and challenges

3.7.1 Pilot testing

Although there is no such thing as 'bug free software', the testing phase of a product in development is essential for generating a high quality end result. This is why it is important that the investigator is checking both the questionnaire and online maps thoroughly. Additionally, pilot tests have been performed by three friends; one is a student in the faculty of geosciences, the other two have different a study background with less map reading experience. The final adjustments were made to the questionnaire and usability test structure based on the pilot tests:

- How to briefly explain the goal and structure of the test, in order to devote most of the time to the main test.

- One of the pilot participants wanted to move the targets to turn the map instead of moving the iPad. This is will not be allowed during the main test because only moving the iPad is used for the navigation control only. Otherwise, the usability analysis becomes too complex.

- Minor spelling mistakes in the questionnaire form were corrected.

3.7.2 Facilities and invitations

Other than refining the usability test itself, the management of the participants and rooms are the final preparations. Rooms were reserved in the Vening Meinesz Building, an Utrecht University building at the Science Park. Students from the faculty of geosciences often have classes or practicals in this building, and therefore this location was chosen as the most optimal for the usability test. The rooms are reserved two weeks in advance for three consecutive days. Although it is expected that most participants will be tested at Utrecht University, the test could take place in any quiet environment, working with mobile devices such as a notebook and an iPad allows for this flexibility.

A week before the scheduled usability tests, the first participants are invited by the investigator. For efficiency, a time schedule was made for assigning the first participants to a time slot based on their availability. Other participants were found by inviting the students on the spot during the test days.

3.7.3 Materials

The following materials are needed for the test: the iPad Pro 12,9" (fully charged), a video camera (charged and extra batteries), an audio recorder, a printed version of the manual, satisfaction interview and MR targets and a pen and paper. The participants are asked to only bring their mobile phones, for filling out the mobile questionnaire. Additionally, the following configurations need to be made: The iPad screen is locked in horizontal orientation, in the Safari browser, only two tabs with the map viewer and online questionnaire are opened, all other browsing history and cookies are is removed from all devices.

3.7.4 Challenges

Although the online questionnaire and maps have been tested various times, unnoticed flaws could still occur in theory. Therefore, the investigator should be prepared to experience a set of difficulties regarding both the hardware, software and the understanding of the participant.

Firstly, the test becomes impractical without a functioning smartphone, notebook and iPad which are all used by the participant. All devices should be fully charged in the morning, an extra phone charging cable should be brought, the Wi-Fi speed on location need to be tested beforehand and all appendices, satisfaction interview sheets and MR targets have to be printed and placed on the desk. Therefore, all hardware, software and other essential materials should be fully operational and present during the test.

Secondly, not only hardware or software problems are possible, there could be a textual misunderstanding of the test by the participant as well. Both the formulated questions and answers should be clear and concise. As an example, the directions of the flows are documented between brackets with an arrows between the first and second number, to make it easier to read (appendix A1).

Thirdly, the participant could have difficulty with navigating the maps. The familiarization phase is designed to let the participant train with the different map controls, especially because working with the MR technique is expected to be new for most participants. Although participants are given the time and instructions to familiarize with the test devices and techniques, theoretically they could lose their control of the navigation for a few seconds, which might be a factor for delay.

Finally other unforeseen difficulties could occur during the test performance. External influences can affect the behaviour of the participant. As an example, when a cup of coffee falls on the ground, the participant will be distracted which causes delay.

4. Methodology - performance

4.1 Introduction chapter

Providing a detailed explanation of the usability test is not the goal of this chapter. All design aspects of the methodology are listed in chapter 3. Instead, any unforeseen events that occurred during the performance of the usability test are explained in this chapter. The actual results of the test are covered in the next chapter, and a more critical reflection on the shortcomings of the usability test is explained in chapter 7.

4.2 Conditions

4.2.1 Facilities

In general, the conditions of all used facilities were good. Most of the participants were tested at one of the available rooms at Utrecht University, but for some participants it was logistically more practical to perform the test at other locations such as my house. In all situations, it was made sure that there was enough light and that the participant could sit or stand to their liking behind a desk. No hardware or software related problems occurred.

4.2.2 Participants

As a result, 21 people have participated in the usability test. Unfortunately, half of the test answers were not saved to the server for one participant, and therefore participant ID 15 is not included in the analysis of the results. With another participant, the questionnaire survey quitted after the third question by accident. The participant was asked to start over again. Only the first 3 maps were affected by this error, but remaining of the results are still valid. Therefore the results of this participant can be used for the analysis. Therefore, the total sample size is 20, which is more than the expected minimum of 16, but not enough for statistical analysis.

As soon as the first 16 participants were covered, the intention was to continue with the ID numbering from ID17 and onwards. However, because there were only a small number of participants remaining, it was decided to continue with ID28 because some difficulties were experiences with ID12, so therefore ID28 can be used as a backup with the identical ordering. There were no problems with ID12 in the end, but an unexpected server problem occurred for ID15 as explained above. Unfortunately, there was no participant with ID31, otherwise the same ordering could have been applied as a backup for ID15.

In general, participants took more time for the first questions of the test because there were not completely ready to start (they clicked on continue while they still wanted to ask a question). This behaviour has caused delay in the first question and the recordings. This is caused because many participants clicked on ‘continue’ too early, while they still had a question to ask to the investigator as part of the familiarization phase. Although the investigator did watch all participants closely, this behaviour was sometimes difficult to prevent because people accidentally continued without realizing that the test would start on the next screen. The same can be applied to the questions of the first map after the screen with the explanation to switch between the map display environments.

Observed how some participants were very descriptive in using the think aloud method, while others remained silent for most of the time although the investigator repeatedly asked to express their thoughts verbally. This difference can be explained as a personal characteristic. It shows how the think aloud would not be sufficient as a method on its own for this research; it needs to be accompanied with other information sources.

4.2.3 Material

As mentioned above, there were no software or hardware related problems. However, some screen recordings are missing because of human failure, instead of a technical failure. This failure was caused because the investigator got distracted by the participant at the beginning of the test, and at the point when it was noticed the first map was already answered. For participants were a recording was not available, it was only for one device. In total the screen recordings are missing for four iPad sessions and 4 PC sessions (table 4.1). The missing screen recordings are not a problem for the analysis, because the recordings intended to be a backup of the observations by the investigator and the measured answers and time by the online questionnaire.

Table 4.1: Overview of the available recordings per participant ID. ✓ = available; X = not available.

ID	iPad screen	PC screen	video	audio
1	X	✓	✓	✓
2	✓	✓	✓	✓
3	✓	✓	✓	✓
4	✓	X	✓	✓
5	✓	✓	✓	✓
6	✓	✓	✓	✓
7	✓	✓	✓	✓
8	✓	X	✓	✓
9	✓	✓	✓	✓
10	✓	X	✓	✓
11	✓	✓	✓	✓
12	✓	✓	✓	✓
13	X	✓	✓	✓
14	✓	✓	✓	✓
15	✓	X	✓	✓
16	✓	✓	✓	✓
17	X	✓	✓	✓
18	✓	✓	✓	✓
28	✓	✓	✓	✓
29	✓	✓	✓	✓
30	X	✓	✓	✓

4.2.4 Questionnaire design

Although the online questionnaire and maps were tested prior to the start of the usability test, minor mistakes were noticed during the performance of the test. As an example, ID1 noticed how there was a question asking to compare between municipalities in map AF1. Instead of 19, the number should have been 9. The participant was informed and continued within 10 seconds with the test, but this has caused confusion and delay. After the end of this test, the investigator informed Yuhang, who directly implemented the change.

Two similar errors in the formulation of the questions were found during the test performance of ID2:

- Map AG4 there is a mistake in one of the answers of the first question (the order is flipped, only one direction is possible per map)
- Map AG2 more answers are possible for the third question: "*Name two of the municipalities with...*"

These errors are similar in nature of the first error: it causes confusion and delay for the participant, but the investigator noticed and explained the questions accordingly. Just as for the first error, Yuhang was able to implement to correct change immediately, so for all upcoming tests this error would not be present.

5. Results

5.1 Introduction chapter

All collected information during the usability test is presented in this chapter. Since a qualitative research method is used, the collected participant observations are combined with the usability test results as a comparison between the four map stimuli, the three usability characteristics and the users' behaviour. For measuring the effectiveness and efficiency, the questionnaire answers, the answer time and the think aloud observations from the main usability test are separately combined for both of the display environments and flow line symbologies. The user satisfaction is based on the results from the satisfaction interview. Because of the small sample size, no statistical tools can be applied to the data. Therefore, all graphs in this chapter are used for descriptive and clarification purposes only.

A think aloud method is used to let the participants speak out their observations and reasonings while interacting with the maps, to support their answers. The total collected information is consists of questionnaire answers, answer time, screen recordings, video recording, audio recording and notes of the observations. For comparing the opinions of the participants, the notes and additional audio recordings are used in the paragraph *Participant observations*. The results of the measured usability characteristics (efficiency, effectiveness and satisfaction) are covered in the paragraph *Participant performance*. Finally, the level of experience with the used techniques and the average duration time are covered in the final paragraph; *Participant experience*. It should be noted that because the participants are presented with the terms AR instead of MR, the term AR is used in this chapter in the way they have been presented to the participants.

5.2 Participant observation

It should be mentioned that some participants spoke much more than others, although the investigator repeatedly asked for their verbal input when a participant was more silent. However the amount of spoken thoughts is a personal characteristic which is something that the investigator can not influence. A comparison between the pro's and con's of both display environments and flow line symbologies can be concluded based on this information.

5.2.1 Display environments

One significant difference of the two display environments is the operation method. The participants complained mainly about the interaction. One of the main differences between the display environments is the navigation control. The majority of the complaints were about the interaction with the maps. Table 5.1 shows an overview of the pros and cons of the display environments, including the corresponding comments made by the participants.

Table 5.1. Pros and cons of the display environments, based on the participant's comments

<table>
<tr><th></th><th>PC screen</th><th>Tablet AR</th></tr>
<tr><td>**Pros**</td><td>**1. Interaction with computer cost less effort.**
"Navigation is easier because I am much more used to using a mouse."

"PC is much easier to navigate."

2. Instead of AR which takes in 'real world' background, the Computer 3D provide users with a cleaner environment to focus on flow maps.
"It is nice that there is a uniformly coloured background for the PC maps."</td><td>**1. Users like the freedom of controlling views with the tablet.**

"I would have preferred to be able to hold the iPad at chest level, because that would make it easier to watch."

"The nice thing about the iPad is that you can hold is vertically, and see the height differences in one instance from a side view."</td></tr>
<tr><td>**Cons**</td><td>**1. Not all users are satisfied with map interaction on the PC. Some participants expect a button to reset the view to the central view.**

"The maps were easier on the PC, but the navigation with the mouse orientation was not ideal."

"The controls are very sensitive, I wish it wouldn't turn all the way upside down."

One user used ctrl-scroll to zoom in on the display, which is not a functionality built in for the map but a general Windows functionality. This indicates the need to watch certain flow line (intersections) more closely, because it is not sufficiently clear to notice this level of detail without the zooming.</td><td>**It takes more efforts to interact with AR, both mentally and physically.**

"The iPad is heavy."

"It is hard to handle both the iPad and questionnaire at the same time."

"You have to perform more operational steps with the iPad because you have to view from any angle, and sometimes stand up or sit down."

"The height is easier to read on the iPad, but the general iPad navigation is annoying. In general, the map reading is clear. But sometimes you need to perform some extra operations to find your answer."

2. Users move a lot in order to change perspectives. But the movement takes more time than simply rotating the maps on the PC.
"You have to perform more operational steps with the iPad because you have to view from multiple angles, and sometimes stand up or sit down."

3. Doubts from users about the significance of AR
"AR maps are very exciting, but it does not add value to the communication of the information in my opinion."

4. Too much freedom of changing views may not always be beneficial.
Some participants tried to work in portrait mode (holding the iPad vertically).

Some participants tried to move the AR targets with one hand, instead of moving the iPad. The investigator kindly asked not to move the target because this is not allowed for this research.

5. Influence of the environments
AR relies on camera to provide its augmentation'. Overexposure or underexposure of the camera will negatively affect the AR map performance.</td></tr>
</table>

Based on the findings on the display environments, it can be concluded that the tablet AR environment is less efficient than the PC screen environment. The first reason is because participants explained they were more experienced and familiar with the interaction of with a keyboard and mouse. The tablet AR requires active movement from the user, but this was perceived as less comfortable after a few minutes in the AR section of the test. It should be noted that the navigation on the PC maps is not ideal likewise. The addition of a 'reset'-button is an example of a feature that can be integrated for the PC environment, but this is not (yet) possible for AR. Since the focus of this research is to compare solely between the four map stimuli, all other features and elements in the maps are similar. Therefore, examples such as the 'reset'-button are not included in the PC map for this usability test. Despite the fact that AR is less efficient in this research, there is potential interest for a new interaction method to view geographical data.

Secondly, a suggestion for further research on AR flow maps is to free the user's hands by fixing the position of the tablet. For this research, the AR marks are fixed on the table. Therefore the users are moving the iPad instead of moving the marks. When the user is free to move the marks with a fixed display, the user does not need to change position. Interacting by using your hands can be regarded as a more human intrinsic method. However, a fixed display resembles the VR technique more than AR. This implies that more insight is needed on the usability of maps displayed in either VR compared to AR.

5.2.2 Flow line symbologies

The comparison between the flow line symbologies is more nuanced than for the previous two map stimuli. Based on the commentary received from the participants during the test, their preferences for height or volume varies. By comparing the usability characteristics with the participant's commentary, the most desirable preference in terms of usability can be concluded. Table 5.2 shows an overview of the pros and cons of the flow line symbologies, including the corresponding comments made by the participants.

It can be concluded that changing the viewpoint is necessary for interpreting the height. Flows with a different height in the AR environments are more difficult to interpret because of the time consuming AR interaction (known from section 1.1). Part of the participants think the height representation reduce visual clutter under certain circumstances; when the focused municipality is in the centre of the map, flows with height variation are easier to distinguish.

For the volume representation, less interaction is needed for the interpretation. This implies that the volume maps have a better usability. But due to the perspective deformation when viewing the flows in 3D, some participants hesitated, or had trouble differentiating between classes 2 and 3, the lowest and highest classes (1 and 4) were considered to be easier to differentiate.

The main factors influencing the performance of the representations are interaction and deformation. To interpret the height, the users need more interaction, whereas the representation of volume is easily influenced by deformation. Therefore, the difference of classes could be enlarged, for preventing deformation. Secondly, the height can reduce visual clutter when the focused municipality is located in the centre. Therefore, changing the height of flows encoded by volume could be another design option, but its usability needs to be further tested.

Table 5.2. Pros and cons of two display environments, based on the participant's comments

	Height	Volume
Pros	**When focused municipality located in the centre, there is less visual clutter on flows with height encoding magnitude.** *"The height maps are working better, and are not as crowded to the 'focus municipality'. Height is way better than the volume for AR."* **In AR environment, participants think visual clutter is less in map with height encoding magnitude** *"The height is way easier for AR. Especially for the question "which one is larger?"*	**Reading flow magnitude from volume requires less efforts. You do not need to change perspective frequently.** *"Line volume is better for comparing between the flows."* *"It is nice to be able to view the maps from above for comparing the volumes."* *"Line volume maps are easier to read"* *"You do not have to move the iPad as much for the maps showing thickness than for the maps showing heights"*
Cons	**To perceive the height, participants need to change perspectives more frequently.** *"The line volume maps can be read within a glance from above, while the height maps require you to move in multiple angles."* *"The maps with height are more difficult to read."* *"Every time it is a matter of changing angles, when you have a map from volume to height."* **Perspective makes it difficult to interpret height class according to legend.** *"The depth effect makes it difficult to differentiate between the classes and compare the flows' heights with the legend."*	**The deformation of perspective makes it difficult to distinguish different classes**. *"Separating class 2 and 3 is very hard for the volume maps."* *"When comparing flows, the volume maps are easier but if you have to tell which is larger, I am not so sure. Even if I change perspective, I am not sure which is larger."*

5.3 Participant performance

There are 20 participants included in the analysis of the results, including thirteen males and seven females. Sixteen participants are aged between 18-25, three participants between 26-30, and one participant is older than 50. Sixteen participants have map reading experience and sixteen participants with 3D maps. Only eleven participants have experience with augmented reality.

There are sixteen maps in total, each map contains three questions which includes one elementary level question and two synoptic level questions. Participants' responding time is recorded automatically, and their answers are scored.

The maps are classified by four groups, and each group contains four maps, divided by the Groningen/Friesland data. Group 1 and group 2 are both displayed on tablet (iPad). Group 1 represent flow magnitude with height and group 2 represent flow magnitude with volume. Group 3 and group 4 are both displayed on computer screen. Group 3 represents flow magnitude with height and group 4 represent flow magnitude with volume.

5.3.1 Efficiency

The answer time of each participant is recorded automatically by the online questionnaire. The time recording starts when opening the question sheet for the first map, which displays three questions for the corresponding map. The recording ends when the last question is answered. Figure 5.1 is a box plot of the time records, showing the variety and the average time spent for each group.

Figure 5.1. Boxplot of the time duration of all maps, divided in four groups. Made by Yuhang Gu

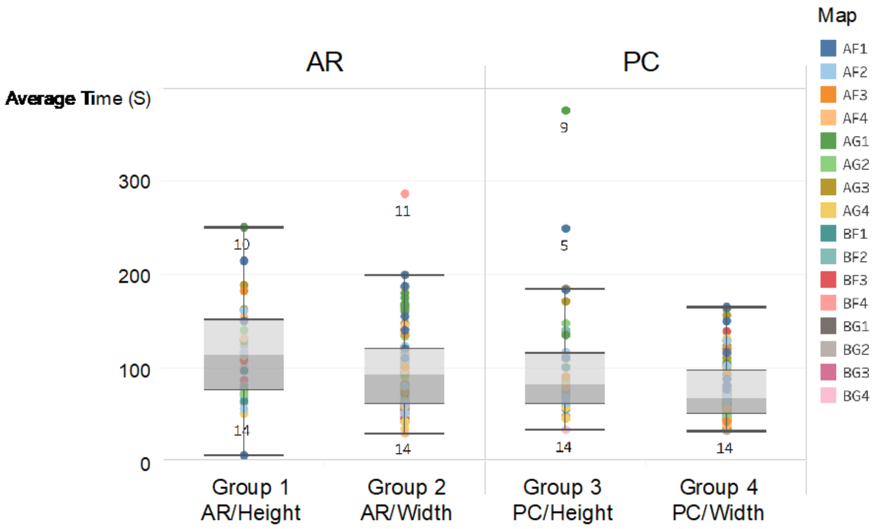


Figure 5.1 supports the conclusion made from the participant comments; flow maps on the PC (group 3 and group 4) are more efficient than flow maps in AR (group 1 and group 2). The representation of volume if more efficient than height. Flows represented with height are answered with the longest duration time. These findings based on the efficiency are in line with the previous results, because the interpretation of height requires more interaction than for volume, while the AR interaction requires more effort than for the PC.

The ordering of the display environments could have affected the results, because their answer speed increased as the test progressed because they became more familiar with the questionnaire and the maps. However, the PC questions were answered more quickly in general, but when the user ends with maps which are more readable this could be experienced as a relief, and could make the previous maps seem more difficult.

5.3.2 Effectiveness

The answers are scored based on the following rules:

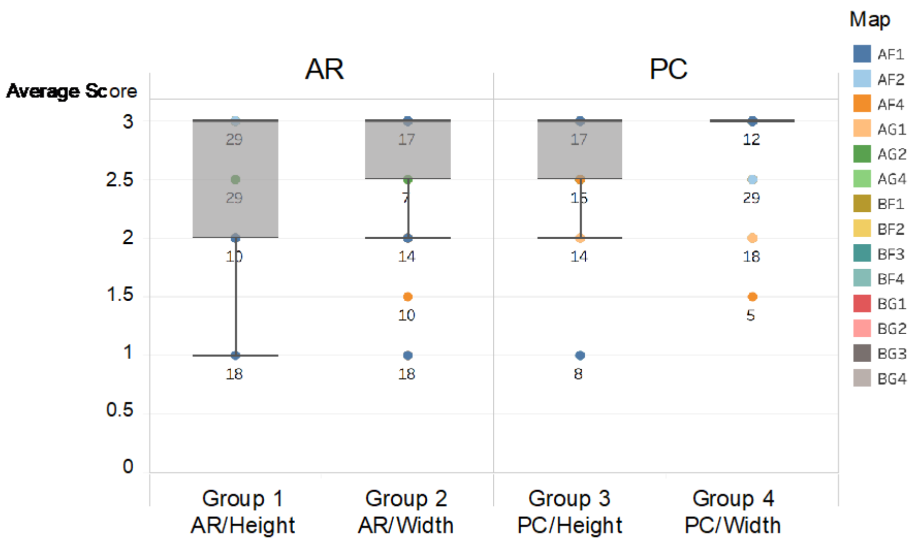
Question 1: For the multiple choice answer, only one answer is correct. When the participant makes the right choice, they receive 1 point.

Question 2: the same as question 1.

Question 3: This question requires participants to find one or two flows. When the correct answer contains 1 flow, the rule is the same as question 1 and question 2. When the answer contains multiple flows, the participant will scored 0.5 point for each correct flow with a maximum score of 1.

Figure 5.2 shows that participants who are using flow maps on the PC have a better performance than those working in AR. For the flow line symbologies, the results for volume are more accurate than for height. The differences are more extreme for the effectiveness scores than for the efficiency measurements. These findings are supporting the conclusion made above..

Figure 5.2 Boxplot of scores of each map, divided in four groups. Made by Yuhang Gu



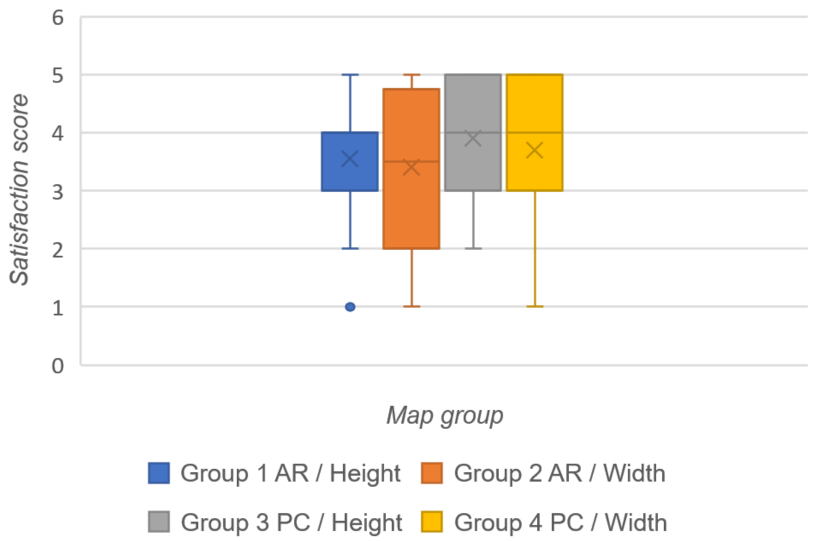
5.3.3 Satisfaction

The satisfaction is measured by the interview at the end of the usability test. Eight of the maps used in the usability test were used for this interview. The same maps were shown to all participants, in the same display environment as they have seen it during the main test. Because participants were presented with the same maps in different environments, the sample is divided into two groups; the odd and even numbered participants. Eleven participants are assigned to group 1 (odd ID) and nine participants to group 2 (even ID). Based on the following statement: *'Map # has a good readability and no signs of visual clutter'*, the participants were asked to assign scores on the satisfaction of the maps ranging from 'strongly disagree' (1) to 'strongly agree' (5).

For the odd participants, the maps starting with ('A') were displayed on the PC, and ('B') in AR, and vice versa for the even participants (see figure X for the sequencing of all maps per user ID). Overall, the participants provided a high satisfaction score with an mean value of 3,59 (3,68 for even; 3,51 for odd). The satisfaction scores of participant ID 11 is ignored for satisfaction analysis, because this participant assigned the highest score of 5 to all maps without providing an explanation. This scoring is perceived as an outlier from the collected data sample, which will not contribute to a clear result. The remaining test data of ID 11 had no peculiarities, and has therefore been used for the other results.

Figure 5.3 shows the satisfaction box plots, grouped similarly as for the efficiency and effectiveness graphs. There are no strong differences in the satisfaction rating of the maps. However, the maps in the AR environment (group 1 and 2) have a lower mean satisfaction rating than for the PC environment (group 3 and 4). It is more difficult to make a conclusion on the preference for the flow line symbology, but the volume has a lower mean score as well as more lower scores than for the maps represented by height. The satisfaction findings are in line with all previous results.

Figure 5.3: **B**oxplot of the satisfaction scores of eight of the total maos, divided in four groups (made by author)



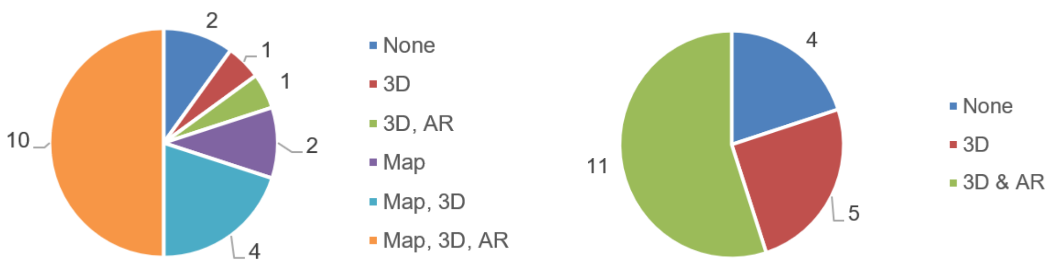
Based on the additional comments made in the satisfaction form, participants mainly based their scores on the maps design, including the geographical spread of the flow lines. For instance, when the focus municipality is located at the edge of the area, the map is perceived as more readable. Four participants commented that the blue lines were better readable than the red lines. Finally, when only a small number of the total number of flows belongs to the smallest or lowest class, the map is better readable. Readability differences caused by a geographical spread of the flows cannot be adjusted, since this caused by the data source.

The goal of this thesis is to minimize the level of visual clutter for all flow maps. Therefore the combination of the satisfaction-, efficiency- and effectiveness scores should indicate whether the preference of the participant is in line with the usability. It should be noted that in general, a positive satisfaction score was given although the participants were critical in their commentary during the main test. It could be regarded as a personal characteristic to be more optimistic when critiquing the work of others. Therefore, the results from the efficiency and effectiveness are regarded as more valuable than the satisfaction interview.

5.4 Participant experience

Amongst the twenty participants, four people claim to have no experience with using maps, one has experience with 3D maps and one has experience with 3D and AR maps (Figure 5.4a). The general 'map experience' option is ignored for the following graphs because the 3D and AR display environments are the essential characteristics for this research. The 20 participants are divided in three groups (figure 3-b): experience of both 3D and AR (11 persons), experience of on 3D (5 persons), and experience of none (4 persons).

Figure 5.4. a) Users experience with maps, 3D maps and AR environment.
b) Sub groups based on participants' experience with 3D and AR (made by author)



5.4.1 Efficiency based on experience

When the average time duration is divided in the three groups based on the map reading experience, there does not seem to exist a prominent difference (figure 5.5). However, when analysed separately, a difference was observed. Figure 5.6a shows difference between participants who have AR experience (red bars) and those not (blue bars). The left part indicates that people with AR experience spend more time on exploring those maps. Figure 5.6b shows the difference between participants with 3D experience and those without. The right part shows that participants with 3D map reading experience spend a little more time than who did not used 3D maps.

Figure 5.5: Average time spend on maps in AR (left) and screen (right, Made by Yuhang Gu

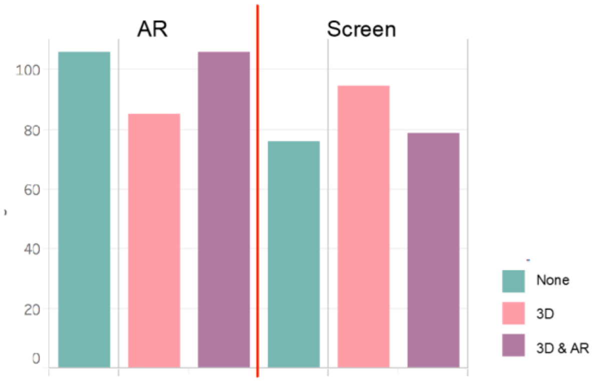
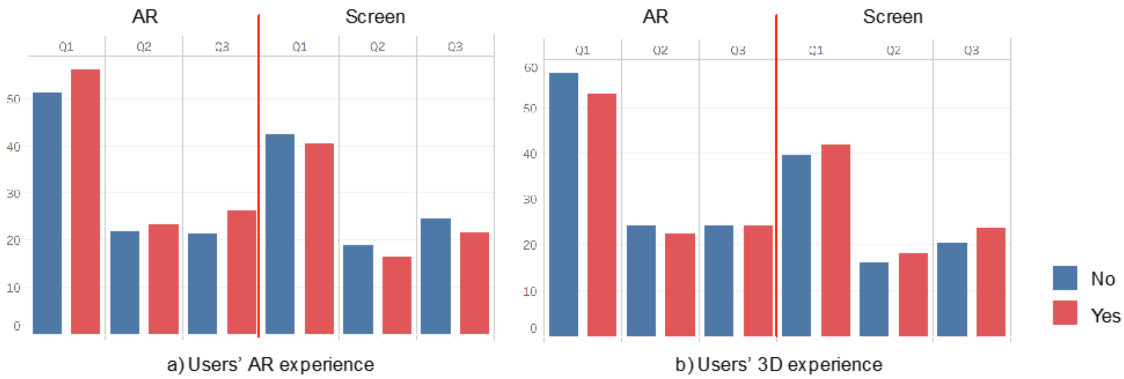


Figure 5.6: Average time spend on each question
a) Time spent on each question from participants with AR experience (red) and no AR experience (blue)
b) Time spent on each question from participants with 3D experience (red) and no 3D experience (blue). Made by Yuhang Gu



5.4.2 Effectiveness based on experience

There is no prominent difference between the three groups based on the map reading experience (figure 5.7). Figure 5.8 shows the same scores, but divided by either the 3D or AR experience. It can be stated that people with experience with AR maps have higher scores than those without experience. Likewise, having 3D map reading experience can help have a better understanding of the maps in the AR environment (5.8).

Figure 5.7: The mean scores for one map grouped by map reading experience.

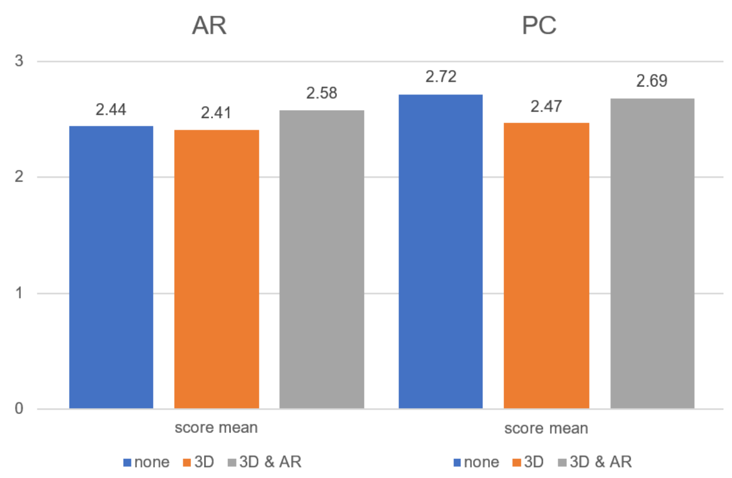
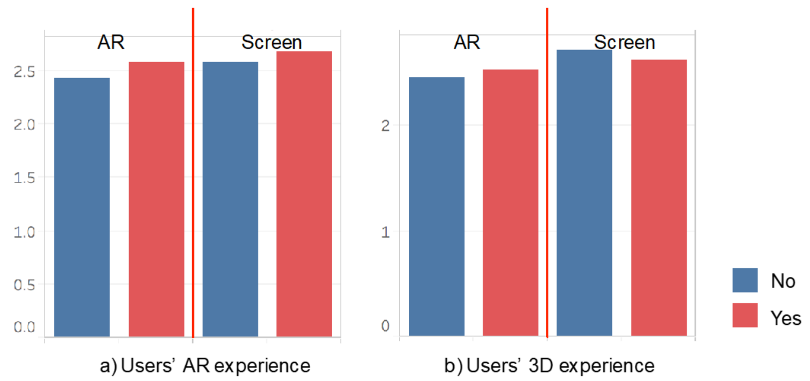


Figure 5.8: Scores of participants with different experience.
a) Time spent on each question from participants with AR experience (red) and no AR experience (blue).
b) Time spent on each question from participants with 3D experience (red) and no 3D experience (blue).



6. Conclusion

The main goal of this research is to investigate to which extent augmented reality, as a modern visualization technique, can be applied to reduce the level of visual clutter in dynamic 3D flow maps. For this thesis, a usability test is performed as a qualitative method for measuring the flow map usability by comparing the efficiency, effectiveness and satisfaction with the 'spoken thought'-comments provided by the participants. The map design is focussed on four map stimuli; augmented reality vs. PC and arc height vs. line volume difference. The test results on the usability are interpreted as means of measuring visual clutter in flow maps.

When considering the efficiency, the flow maps displayed on the PC were answered more quickly than those in MR. This delay is caused because navigating in the MR environment requires the user to actively move more than for the maps on the PC. The maps represented by line volume variation are answered with a shorter duration time than for the maps styled with height. For the effectiveness, the results are the same as for the efficiency, but the differences between the map stimuli are stronger. Finally, the satisfaction outcomes indicate a generally positive satisfaction rating (3,59 out of 5) but also the smallest differences between the map stimuli. All three usability characteristics indicate the same outcomes; a better usability is achieved for flow maps displayed in MR instead of the PC environment and for the line volume instead of the arc height representation. An MR environment displayed on a tablet is perceived as a less comfortable navigation tool compared to a PC's mouse and keyboard to display 3D flow maps. Additionally, the flow line symbology height is perceived as less comfortable compared to the line volume. This is mainly due to the needs of the participant to view the model styled by height from multiple angles in order to retrieve the information. Although MR can enable new visualization opportunities, the usability of a flow map does not automatically improve when compared to the same 3D map displayed on a PC screen.

The map stimuli are not the only influential factors on the flow map usability. Participants explained how the geographical spread of flow lines on the map are influencing the map usability. Additionally, a high amount of lines located at the centre of the map are perceived as less readable, as well as a high amount of flows classified in the two middle classes, 2 and 3. This indicates how a data source and a geographical spread has effects on the usability. Since data sources and locations are fixed, these entities cannot be modified for the purpose of improving the usability.

It can be assumed that the MR display environment does not improve the flow map usability. However, it is too early to make assumptions about the suitability of the MR environment for a visual clutter reduction on flow maps, because this research only focussed on a small selection of symbology and display environment criteria. Because of the likelihood that techniques such as MR will gradually improve over time, its usability might have a different outcome in the future, but the outcomes of this research indicate that the PC environment is more capable than MR to minimize the effect of visual clutter on flow maps today.

7. Discussion

7.1 Interpretation

Two data gathering components are used in this research; a focus group session with two experts and a usability test with 20 participants. The feedback derived from the focus group session is used to optimize the questionnaire and maps for the final test. After the performance of the usability test, the combination of the results on the usability characteristics (efficiency, effectiveness and satisfaction) and the comments made by the participants are the final results on the usability of the flow maps. The individual results of the usability characteristics and the additional observations are equal.

Based on the outcomes of all the usability measures, the MR display environment is rated as a lower usability than the PC display environment and the flow line symbology of arc height was rated lower than line volume. This is not in line with the hypothesis that the MR environment, with its more immersive and interactive quality, would enhance the usability of the flow maps. Some participants explained that they were more familiar with the 'mouse and keyboard' controls of the PC, that the iPad was too heavy for the usage of an extended period of time or that the navigation with the iPad requires more effort and therefore more time. This interpretation should be treated carefully because the scope of this research did not enable a quantitative methodology. Although the MR environment scored a lower usability based on the conditions in this usability test, there is room for improvement when using different flow line symbologies, hardware or other technical measures in further research.

It should be mentioned that this result could partly be caused because of the participant's lesser experience with MR compared to the PC, even though they were familiarized during the start of the test. The difference in the usability results between the two environments are consistent. Further research with a larger sample size and a quantitative methodology should point out whether these usability differences are significant or not.

The think aloud method enabled the participants to give additional information along with the questionnaire answers on the four tested map stimuli. Additional findings are the preference to read blue coloured flows over the red flows and the flows that belong to class 1 and 4 were easier to distinguish than those in class 2 and 3. The data spread also seems to be an important factor, since people preferred maps with the flows clustering near the edge of the map area instead of in the centre. These findings are not part of the scope and the main results, but are worth mentioning because multiple people came up with the same comments.

This research adds to on the study by Jenny, who listed a set of cartographic styling measures for reducing visual clutter on static 2D flow maps (et.al., 2016). By visualizing these as dynamic 3D flow maps, there are more visualization and navigation possibilities which could provide chances for a more usable map with less visual clutter. This research aimed to fill the gap on how visual clutter in dynamic 3D flow maps could be reduced. Instead of measuring visual clutter directly, the usability of the end product, in this case the flow maps, was measured as an indirect measure for visual clutter.

7.2 Limitations

Right before the first map of main test appears, the user needs to click ‘continue’ on an initial screen of the display environments section, MR or PC. Although all participants were asked in advance to wait for the approval before clicking continue, a small number of participants forgot or did not notice that they landed on this specific screen and continued too soon. Therefore, some delay could have been caused for the first maps if all people clicked ‘continue’ when all explanations and recording preparations were finished. The same problem occurred halfway in the usability test when switching between display environments. In those situations, all recording materials still had to be turned on, while the automated timer from the online questionnaire started recording already. This resulted in missing recording data for the first few seconds and an invalid efficiency score for the first map of the specific display environment. Despite this, due to quick acting of the researcher, the extra time recorded was limited to a few seconds per instance.

Not all participant recordings have been collected. Out of the eight missing instances, either a recording did not work because of a memory issue or the investigator was distracted by the questions asked by the participant. This caused no problem, because all recordings were served as backup material, and a maximum of one recording was missing for each of the eight instances per participant. However, a future usability test could be improved by spending more time on performing pilot tests to minimize these errors even further.

Although the questionnaire and the online maps are tested comprehensively, minor textual mistakes did occur at the start of the usability test. As soon as the specific test was finished, Yuhang Gu was contacted and managed to update and change the questionnaire immediately before the second participant started. This is not an ideal situation, since all participants should have been presented with the exact same usability test. However, the textual mistakes were corrected and one label number which overlapped the legend was verbally corrected by the investigator when this map was opened.

Another weakness related to the map legend is the missing explanation of the flow line directionality in the map view. The meanings of the blue and red coloured lines are given in the text presented in the questionnaire, but as soon as the participant is looking at the maps and answering the questions, this information is no longer visible. At times when the participant asked for the direction of the colours, the investigator assisted by saying the direction linked per colour. When a legend with the blue and red colouring was given for each map, this could have been prevented.

7.3 Recommendations

Based on the results and limitations of this research, the following recommendations are listed for further usability research or in combination with the reduction of visual clutter for 3D flow maps. Firstly, the methodology of this thesis can be applied for a follow up research on the reduction of visual clutter in 3D flow maps. Instead of using a qualitative method, a quantitative approach is recommended. When a higher sample size of roughly 50 participants or more would be accomplished, the correlations found in the results could be tested on its statistical significance. This would enable more insight in the correlations between the map stimuli and the usability, that visual clutter could be reduced on flow maps. When working with a larger sample group, there could be differences between people with map reading experience and without, for analysing whether there are significant differences in the usability results. This is only beneficial, when the produced flow maps are intended to be used by a wide variety of end users as its audience.

Secondly, further usability research can be performed on interactive flow maps. Interactive flow maps are more challenging to develop for an MR environment, but could enhance the overall usability. When the user is able to point to or click on an individual data feature, additional attribute information can be presented through a pop up. This addition would enable a more data rich flow map visualizations in the MR environment. Another example is the integration of animated features, such as flow lines as a symbology characteristic. However, the implementation of this technique requires more technical expertise. Another example of a more interactive flow map could be the implementation of animated flow lines to represent time.

Thirdly, the usage of the MR environment could be critically reflected upon for the purpose of a flow map. The tablet would be mounted onto the table in a fixed position aiming on the target, the participant's hands would be free from the tablet, so they can turn the target and hold their phone for the questionnaire instead. This could enhance the comfort, but the MR technology is limited and it becomes more similar to the navigation controls of VR glasses since the hands of the user are free from the display.

Other than revising the application of the MR technique, maps displayed in 2D could be added to the total set of maps to compare the difference in usability, and potentially visual clutter, between static 2D and dynamic 3D maps. When adding one or more elements to a future follow up usability test, the total duration is extended. It is crucial to consider the importance of the focus of the participant for the validity of the results, Instead of extending the total test duration, the total population sample could be divided in groups based on the display environments.

The preparation of a usability test is essential for a high quality of the end results. When the test has begun, no more alterations can be made. Additionally, people who have participated can not be invited for a resit, because their experience with the maps could influence the results. Performing more individual pilot tests with people from both geographical and non-geographical backgrounds prior to the usability test might generate more feedback, which is valuable for finding and solving the final errors in time.

8. References

Adams, A., & Cox, A. L. (2008). Questionnaires, in-depth interviews and focus groups. In *Research Methods for Human Computer Interaction* (pp. 17–34). Cambridge University Press.

Andrienko, Andrienko, Demsar, Dransch, Dykes, Fabrikant, … Tominski. (2010). Space, time and visual analytics. *International Journal of Geographical Information Science*, *24*(10), 1577–1600. https://doi.org/10.1080/13658816.2010.508043

Andrienko, N., & Gennady, A. (2006). Exploratory analysis of spatial and temporal data: A systematic approach. In *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. https://doi.org/10.1007/3-540-31190-4

Bach, B., Dragicevic, P., Archambault, D., Hurter, C., & Carpendale, S. (2017). A Descriptive Framework for Temporal Data Visualizations Based on Generalized Space-Time Cubes. *Computer Graphics Forum*. https://doi.org/10.1111/cgf.12804

Bertin, J. (1967). *Sémiologie graphique: les diagrammes, les réseaux, les cartes.* (G.-V. & Cie., Ed.). Paris: École des hautes études en sciences sociales/Mouton-Gauthier-Villars.

Bevan, N., Carter, J., & Harker, S. (2015). ISO 9241-11 revised: What have we learnt about usability since 1998? In Springer International Publishing (Ed.), *International Conference on Human-Computer Interaction* (pp. 143–151). Retrieved from http://www.springer.com/series/7409

Boeije, H., 't Hart, H., & Hox J. (2009). Onderzoeksmethoden. *Den Haag: Boom Lemma Uitgevers*.

Börner, K. (2015). *Atlas of Knowledge: Anyone Can Map*. 224. https://doi.org/10.1215/01636545-1987-38-115

Cambray, B. de, & Yeh, T.-S. (1994). A Multidimensional (2D, 2.5D, and 3D) Geographical Data Model. *International Conference on Management of Data (COMAD'94)*, (May), 317–336.

Carbonell Carrera, C., & Bermejo Asensio, L. A. (2017). Augmented reality as a digital teaching environment to develop spatial thinking. *Cartography and Geographic Information Science*, *44*(3), 259–270. https://doi.org/10.1080/15230406.2016.1145556

CBS. (2017). Verhuisde personen tussen gemeenten. *Centraal Bureau Voor de Statistiek*. Retrieved from https://opendata.cbs.nl/statline/#/CBS/nl/dataset/60048ned/table?fromstatweb

Claudel, M., Nagel, T., & Ratti, C. (2016). From origins to destinations: The past, present and future of visualizing flow maps. *Built Environment*, *42*(3), 338–355. https://doi.org/10.2148/benv.42.3.338

Cook, K. A., & Thomas, J. J. (2005). *Illuminating the path: The research and development agenda for visual analytics*. Richland, WA (United States): Pacific Northwest National Lab (PNNL).

Crampton, J. W. (2001). Maps as social constructions: Power, communication and visualization. *Progress in Human Geography*, *25*(2), 235–252. https://doi.org/10.1191/030913201678580494

Dibiase, D., Maceachren, A. M., Krygier, J. B., & Reeves, C. (1992). Animation and the Role of Map Design in Scientific Visualization. *Cartography an Geographic Information Systems*, *19*(4), 201–214, 265–266. Retrieved from https://www.geovista.psu.edu/publications/1992/DiBiase_C&GIS_92.pdf

Farr, B. C. (2008). Designing Qualitative Research. *Transformation: An International Journal of Holistic Mission Studies*, *25*(2–3), 165–166. https://doi.org/10.1177/026537880802500310

Field, K., & Saunder, D. (2018). *Cartography*. ESRI Press.

Fisher, P., Dykes, J., & Wood, & J. (1993). The Cartographic Journal The World of Mapping Map design and visualization Map design and visualization. *The Cartographic Journal*, *30*(2), 136–142. https://doi.org/10.1179/000870493787859960

Garlandini, S., & Fabrikant, S. I. (2009). Evaluating the Effectiveness and Efficiency of Visual Variables for Geographic Information Visualization. *Spatial Information Theory*, 195–211. Retrieved from https://link.springer.com/content/pdf/10.1007%2F978-3-642-03832-7.pdf

Gu, Y., Kraak, M.-J., & Engelhardt, Y. (2017). Revisiting flow maps: a classification and a 3D alternative to visual clutter. *Proceedings of the ICA*, *1*(51). Retrieved from https://pdfs.semanticscholar.org/df79/f0a5e88b038e73a9d875ccfa2c3120211d67.pdf

Haeberling, C. (1999). Symbolization in topographic 3D maps: Conceptual aspects for user-oriented design. *19th International Cartographic Conference*, 1037–1044.

Hägerstrand, T. (2004). The two vistas. *Geografiska Annaler, Series B: Human Geography*, *86*(4), 315–323. https://doi.org/10.1111/j.0435-3684.2004.00170.x

Harrie, L., Stigmar, H., & Djordjevic, M. (2015). Analytical Estimation of Map Readability. *ISPRS International Journal of Geo-Information*. https://doi.org/10.3390/ijgi4020418

Harrower, M., & Fabrikant, S. (2008). *The Role of Map Animation for Geographic Visualization*. Retrieved from http://maps.unomaha.

Henrique, G., Pereira, de A., Stock, K., Stamato Delazari, L., & Centeno, J. A. S. (2017). Augmented Reality and Maps: New Possibilities for Engaging with Geographic Data. *Cartographic Journal*, *54*(4), 313–321. https://doi.org/10.1080/00087041.2017.1411417

International Organization for Standardization. (2018). Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts. *Standard No. ISO 9241-11*. Retrieved from https://www.iso.org/standard/63500.html

Jacobsen, N. E., Hertzum, M., & John, B. E. (1998). The Evaluator Effect in Usability Tests. In *CHI* (Vol. 98). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.418.6674&rep=rep1&type=pdf

Jenny, B., Stephen, D. M., Muehlenhaus, I., Marston, B. E., Sharma, R., Zhang, E., & Jenny, H. (2018). Design principles for origin-destination flow maps. *Cartography and Geographic Information Science*, *45*(1), 62–75. https://doi.org/10.1080/15230406.2016.1262280

Johnson, H., & Nelson, E. S. (1998). Using Flow Maps to Visualize Time-Series Data: Comparing the Effectiveness of a Paper Map Series, a Computer Map Series, and Animation. *Cartographic Perspectives*, *0*(30), 47–64.

Keim, D. A., Mansmann, F., Schneidewind, J., Thomas, J., & Ziegler, H. (2008). Visual Analytics: Scope and Challenges. *Visual Data Mining*, 76–90. Retrieved from http://kops.ub.uni-konstanz.de/volltexte/2009/6842

Kim, K., Hwang, J., & zo, H. (2016). Understanding users' continuance intention toward smartphone augmented reality applications. *Information Development*, *32*(2), 161–174. https://doi.org/10.1177/0266666914535119

Koussoulakou, A., & Kraak, M. J. (1992). Spatia-temporal maps and cartographic communication. *The Cartographic Journal*, *29*(2), 101–108. https://doi.org/10.1179/000870492787859745

Kraak, M.-J. (1998). The Cartographic Visualization Process: From Presentation to Exploration. *The Cartographic Journal*, *35*(1), 11–15. https://doi.org/10.1179/000870498787074100

Kraak, M.-J. (2011). Cartography and Geographic Information Science Is There a Need for Neo-

Cartography? *Cartography and Information Science*, *38*(2), 73–78. https://doi.org/10.1559/1523040638273

Kraak, M. J., & Ormeling, F. (2013). Cartography: Visualization of geospatial data, third edition. In *Cartography: Visualization of Geospatial Data, Third Edition* (Vol. 9781315847). https://doi.org/10.4324/9781315847184

Lukačovič, I. (2018). Windy (weather service). Retrieved September 15, 2018, from https://www.windy.com

MacEachren, A. M. (1994). Visualization in Modern Cartography: Setting the Agenda. In *Modern Cartography Series* (Vol. 2, pp. 1–12). https://doi.org/10.1016/B978-0-08-042415-6.50008-9

MacNeish, H. F. (1922). Euler Squares. *The Annals of Mathematics*, *23*(3), 221. https://doi.org/10.2307/1967920

McDonagh-Philp, D., & Bruseberg, A. (2000). Using focus groups to support new product development. *Institute of Engineering Designers Journal*, *26*(5), 4–9.

Milgram, P., & Kishimo, F. (1994). A taxonomy of mixed reality. *IEICE Transactions on Information and Systems*, *77*(12), 1321–1329.

Monmonier, M. (2005). Lying with Maps. *Institute of Mathematical Statistics*, *20*(5), 215–222. Retrieved from https://www.jstor.org/stable/pdf/20061176.pdf

Moro, C., Štromberga, Z., Raikos, A., & Stirling, A. (2017). The effectiveness of virtual and augmented reality in health sciences and medical anatomy. *Anatomical Sciences Education*, *10*(6), 549–559. https://doi.org/10.1002/ase.1696

Nielsen, J., & Landauer, T. K. (1993). A mathematical model of the finding of usability problems. *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems*, 206–213.

Olmsted-Hawala, E. L., Murphy, E. D., Hawala, S., & Ashenfelter, K. T. (2010). Think-aloud protocols: a comparison of three think-aloud protocols for use in testing data-dissemination web sites for usability. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2381–2390. https://doi.org/10.1145/1753326.1753685

Parker, A., & Tritter, J. (2007). Focus group method and methodology: current practice and recent debate. *International Journal of Research & Method in Education, 29*(1), 23–37. https://doi.org/10.1080/01406720500537304

Pred, A. (1977). The Choreography of Existence: Comments on Hägerstrand's Time-Geography and Its. In *Source: Economic Geography* (Vol. 53). Retrieved from https://www-jstor-org.proxy.library.uu.nl/stable/pdf/142726.pdf?refreqid=excelsior%3A408315391a8c956938aa51d9626399e2

Read, J. C. A., Phillipson, G. P., Serrano-Pedraza, I., David Milner, A., & Parker, A. J. (2010). Stereoscopic vision in the absence of the lateral occipital cortex. *PLoS ONE*, *5*(9), 1–14. https://doi.org/10.1371/journal.pone.0012608

Rød, J. K., Ormeling, F., & van Elzakker, C. (2001). An agenda for democratising cartographic visualisation. *Norsk Geografisk Tidsskrift*, *55*(1), 38–41. https://doi.org/10.1080/00291950120034

Romat, H., Appert, C., Bach, B., Henry-Riche, N., & Pietriga, E. (2018). Animated edge textures in node-link diagrams: A design space and initial evaluation. *Conference on Human Factors in Computing Systems - Proceedings*, *2018-April*. https://doi.org/10.1145/3173574.3173761

Roth, R. E. (2017). Visual Variables. In *International Encyclopedia of Geography: People, the Earth,* *Environment and Technology*. https://doi.org/10.1002/9781118786352.wbieg0761

Seipel, S. (2013). Evaluating 2D and 3D geovisualisations for basic spatial assessment Evaluating 2D and 3D geovisualisations for basic spatial assessment. *Behaviour & Information Technology*, *32*(8), 845–858. https://doi.org/10.1080/0144929X.2012.661555

Semmo, A., Trapp, M., Jobst, M., & Döllner, J. (2015). Cartography-Oriented Design of 3D Geospatial Information Visualization-Overview and Techniques. *The Cartographic Journal*, *52*(2), 95–106. https://doi.org/10.1080/00087041.2015.1119462

Stephen, D. M., & Jenny, B. (2017). Automated layout of origin-destination flow maps: U.S. county-to-county migration. *Journal of Maps*, *13*(1), 46–55. https://doi.org/10.1080/17445647.2017.1313788

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*(sup1), 234-240. Retrieved from https://about.jstor.org/terms

Townsend, A. M. (2013). *Smart cities: Big data, civic hackers, and the quest for a new utopia,*. WW Norton & Company.

Tsou, M. H. (2011). Revisiting web cartography in the United States: The rise of user-centered design. *Cartography and Geographic Information Science*, Vol. 38, pp. 250–257. https://doi.org/10.1559/15230406382250

Tussyadiah, I. P., Jung, T. H., & tom Dieck, M. C. (2018). Embodiment of Wearable Augmented Reality Technology in Tourism Experiences. *Journal of Travel Research*. https://doi.org/10.1177/0047287517709090

van Elzakker, C. (2004). The use of maps in the exploration of geographic data. *Utrecht University*.

VPRO. (2018). *Nederland van Boven*. Retrieved from https://www.vpro.nl/programmas/nederland-van-boven/lees/achtergronden/alle-datavisualisaties.html

Wood, J., Dykes, J., & Slingsby, A. (2010). Visualisation of Origins, Destinations and Flows with OD Maps. *The Cartographic Journal*, *47*(2), 117–129. https://doi.org/10.1179/000870410X12658023467367

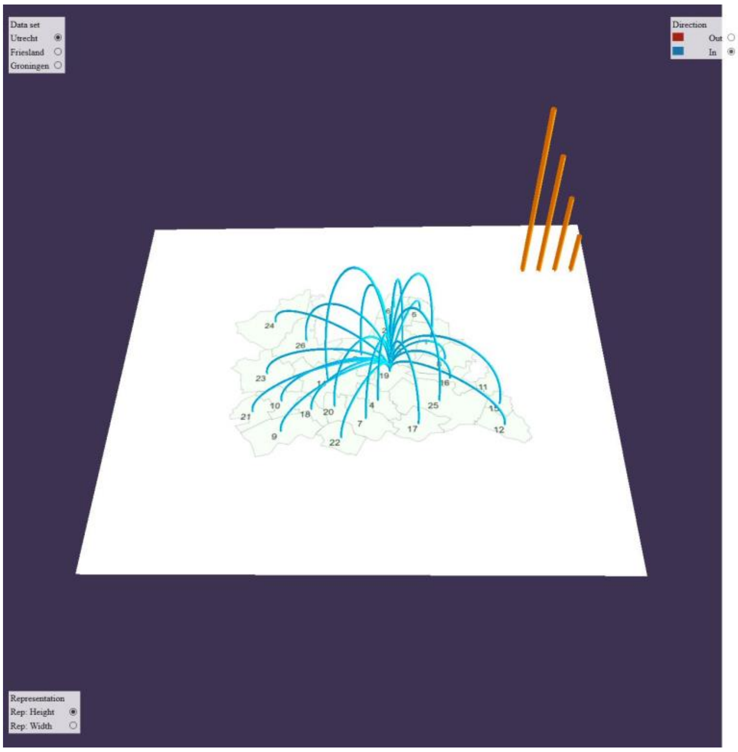
Yang, Y., Jenny, B., Dwyer, T., Marriott, K., Chen, H., & Cordeil, M. (2018). Maps and Globes in Virtual Reality. *Computer Graphics Forum*, *37*(3). https://doi.org/10.1111/cgf.13431

A1. Usability test questionnaire

All questions are listed per map, illustrated with a screenshot.

Familiarization

Map TU1



PC 2.1: When comparing the flows between municipality 19-10 and 19-20, which flow is the smallest?

1. *[The flow from municipality X to municipality X]* **19-10**

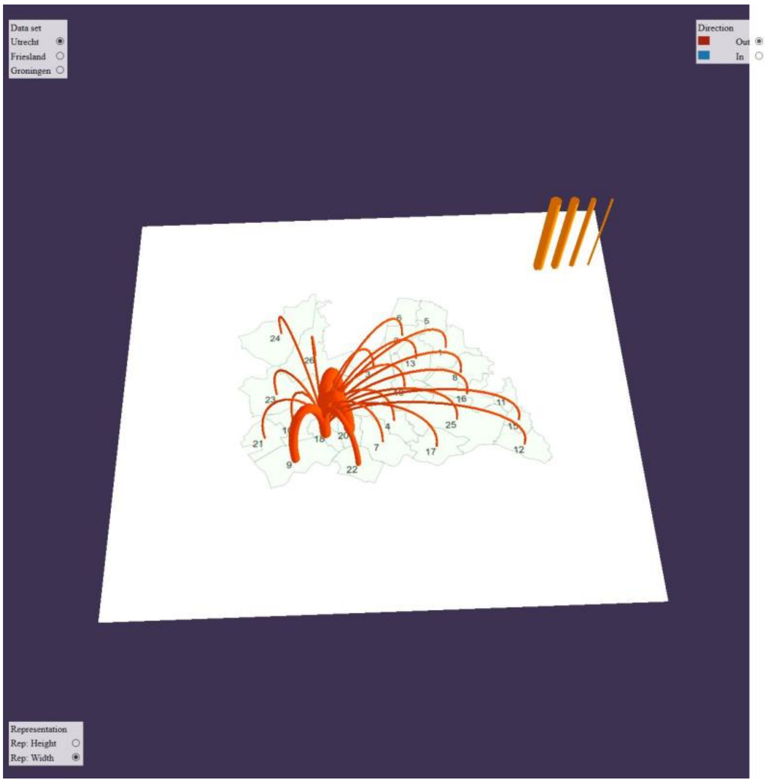
PC 2.1: To which class does the flow from municipality 19 to 15 belong?

1. *class 1*
2. ***class 2***
3. *class 3*
4. *class 4*

M2: What is the origin municipality of the largest flow going to municipality 19?

1. [blank] **14**

Map TU2



PC 1.1: When comparing the flows between municipality 18-20 and 18-22, which flow is the largest?

1. *[The flow from municipality X to municipality X]* **18-20**

PC 1.2: To which class does the flow from municipality 18-26 belong?

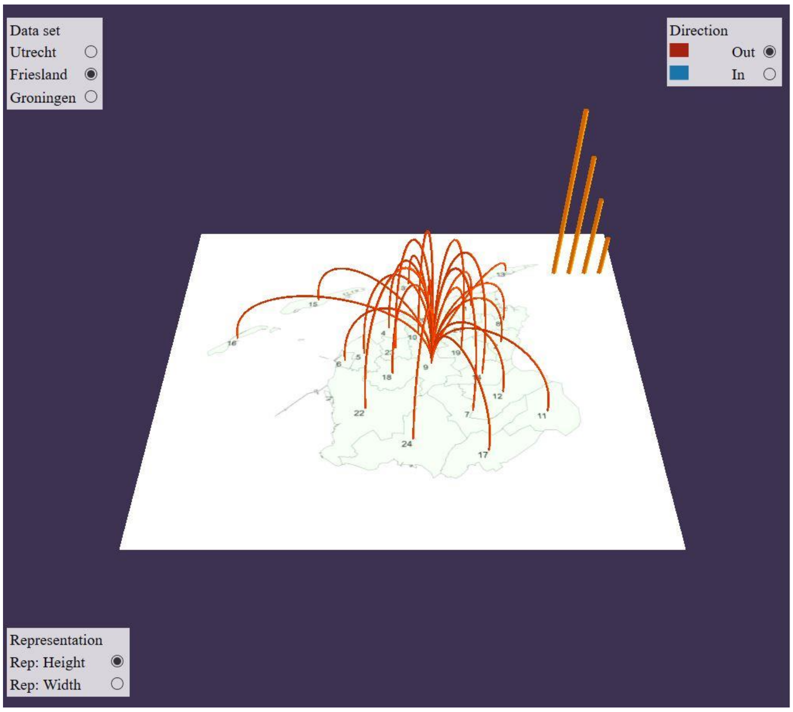
1. ***class 1***
2. class 2
3. class 3
4. class 4

M1: Name the two destination municipalities with the largest flows departing from municipality 18?

1. [blank] **14, 20**

Main usability test

Map 1.AF1



PC 1.1: When comparing the flows between municipality 9-10 and 9-11, which flow is the largest?

1. *[The flow from municipality X to municipality X]* **9-10**

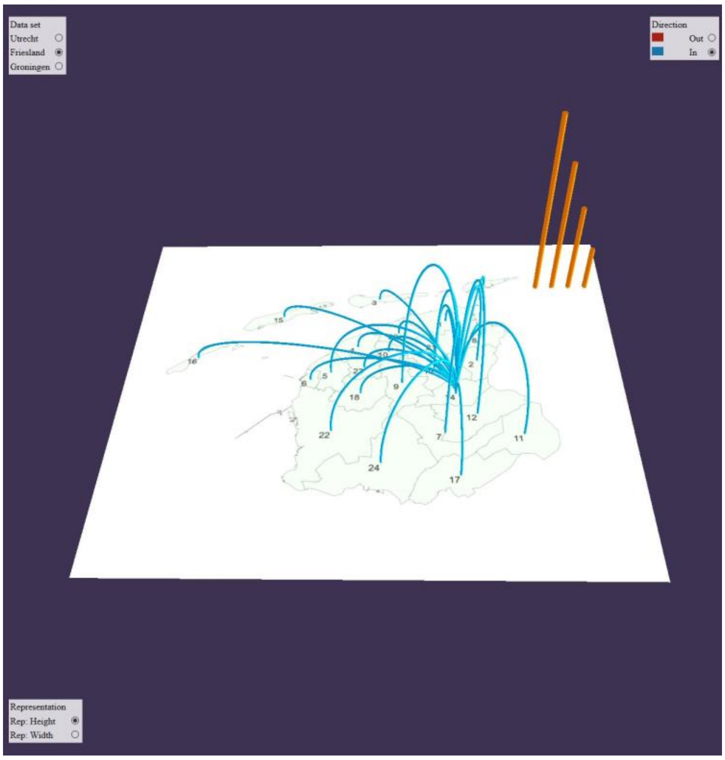
PC 1.2: To which class does the flow from municipality 9-7 belong?

1. *class 1*
2. *class 2*
3. *class 3*
4. ***class 4***

M1: What is the destination municipality of the smallest flow departing from municipality 19?

1. [blank] **13**

Map 2.AF2



PC 2.1**:** When comparing the flows between municipality 14-10 and 14-11, which flow is the largest?

1. [*The flow from municipality X to municipality X*] **14-11**

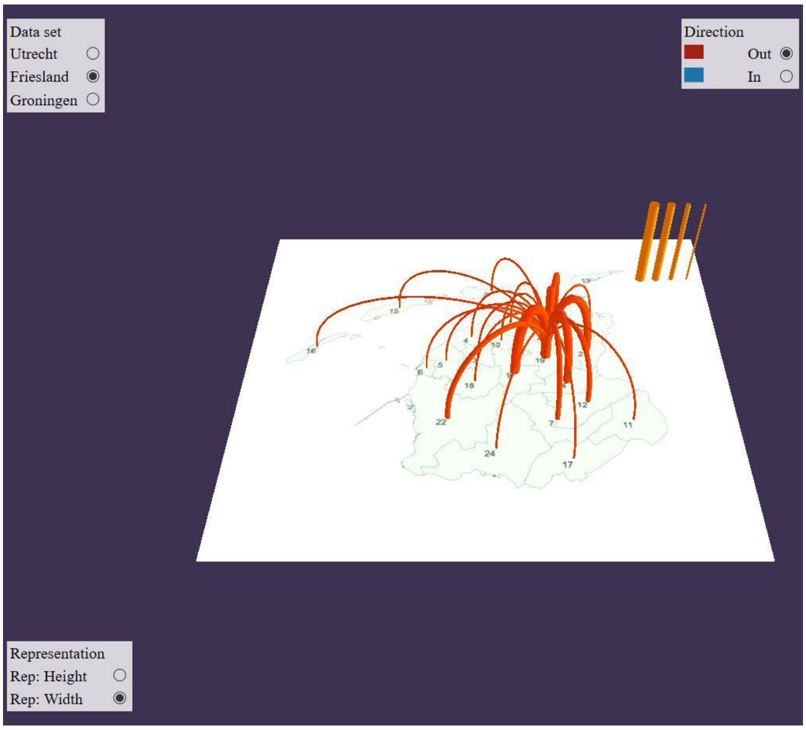
PC 2.2: To which class does the flow from municipality 14-16 belong?

1. ***class 1***
2. *class 2*
3. *class 3*
4. *class 4*

M2: Name the two origin municipalities with the largest flows going to municipality 14.

1. [blank] **9 and 12**

Map 3.AF3



PC 3.1: When comparing the flows between municipality 19-2 and 19-9, which flow is the largest?

1. *[The flow from municipality X to municipality X]* **19-9**

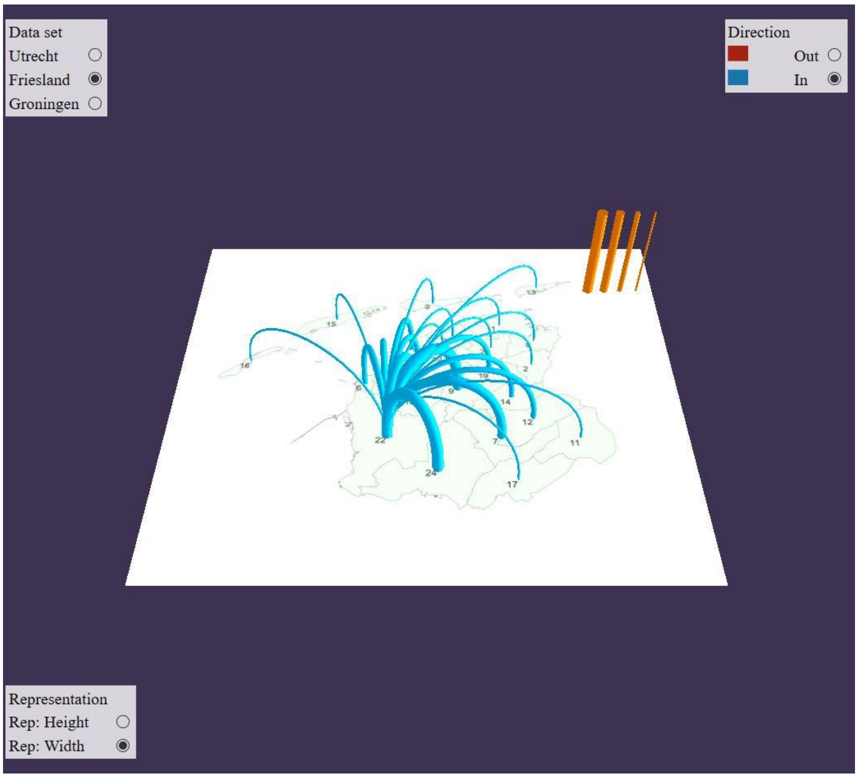
PC 3.2: To which class does the flow from municipality 19-22 belong?

1. *class 1*
2. ***class 2***
3. *class 3*
4. *class 4*

M3: What is the destination municipality of the highest flow departing from municipality 19?

1. [blank] **9**

Map 4.AF4



PC 4.1**:** When comparing the flows between municipality 22-11 and 22-12, which flow is the smallest?

1. *[The flow from municipality X to municipality X]* **22-11**

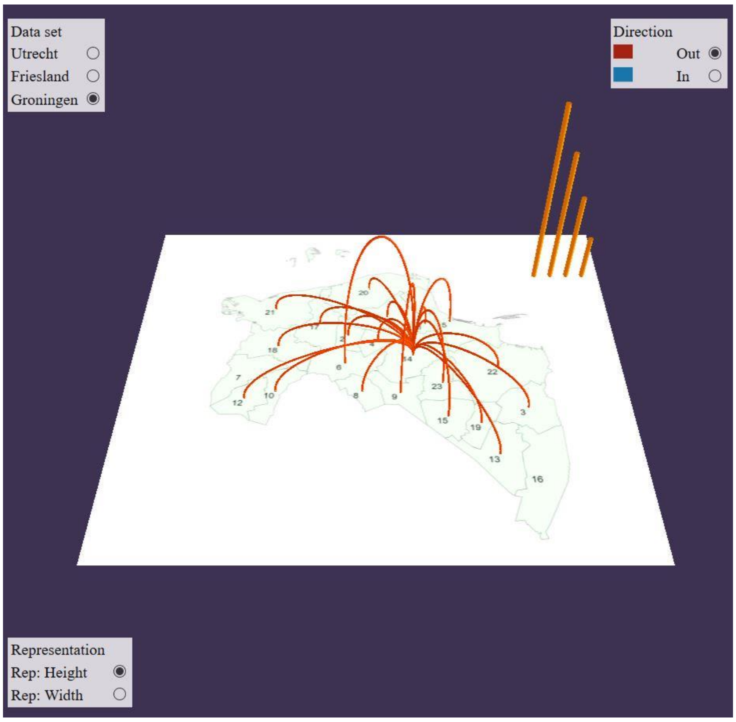
PC 4.2: To which class does the flow from municipality 22 to 24 belong?

1. *class 1*
2. *class 2*
3. *class 3*
4. ***class 4***

M4: Name the two origin municipalities with the largest flows going to municipality 22.

1. [blank] **9, 24**

Map 5.AG1



PC 5.1: When comparing the flows between municipality 14-5 and 14-9, which flow is the largest?

1. *[The flow from municipality X to municipality X]* **14-9**

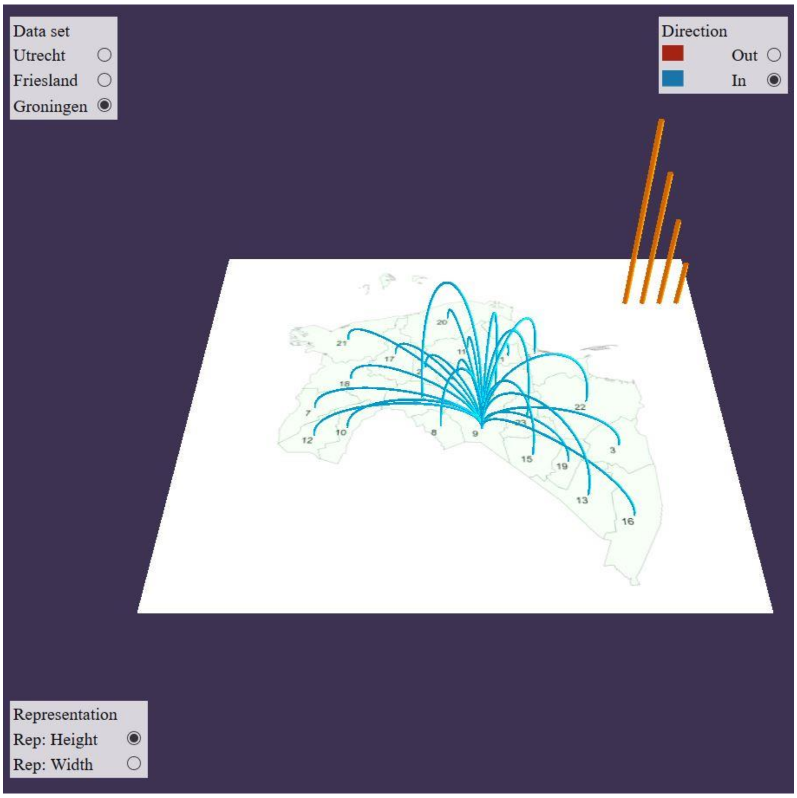
PC 5.2: To which class does the flow from municipality 14-6 belong?

1. *class 1*
2. *class 2*
3. *class 3*
4. ***class 4***

M5: What is the destination municipality of the smallest flow departing from municipality 14?

1. [blank] **6**

Map 6.AG2



PC 6.1**:** When comparing the flows between municipality 9-7 and 9-8, which flow is the largest?

1. [*The flow from municipality X to municipality X*] **9- 8**

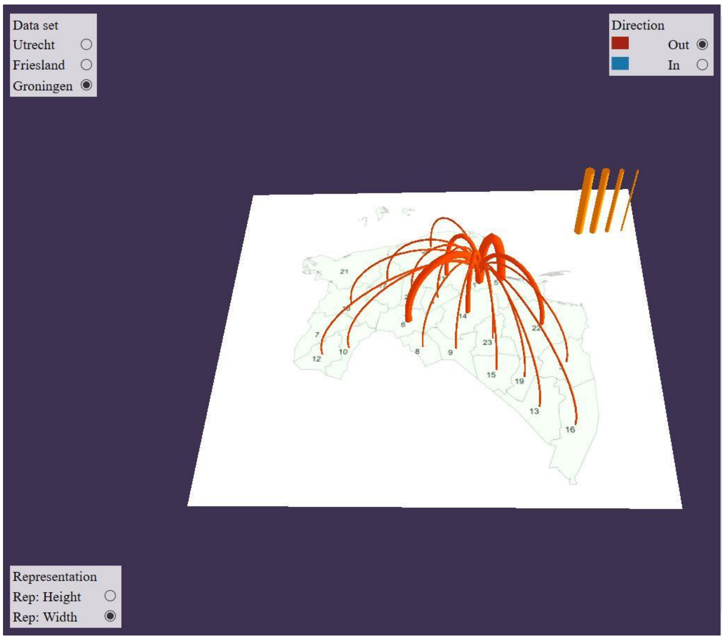
PC 6.2: To which class does the flow from municipality 9-15 belong?

1. *class 1*
2. ***class 2***
3. *class 3*
4. *class 4*

M6: Name the origin municipality with the largest flow going to municipality 9.

1. [blank] **6**

Map 7.AG3



PC 7.1**:** When comparing the flows between municipality 1-22 and 1-23, which flow is the smallest?

1. *[The flow from municipality X to municipality X]* **1-22**

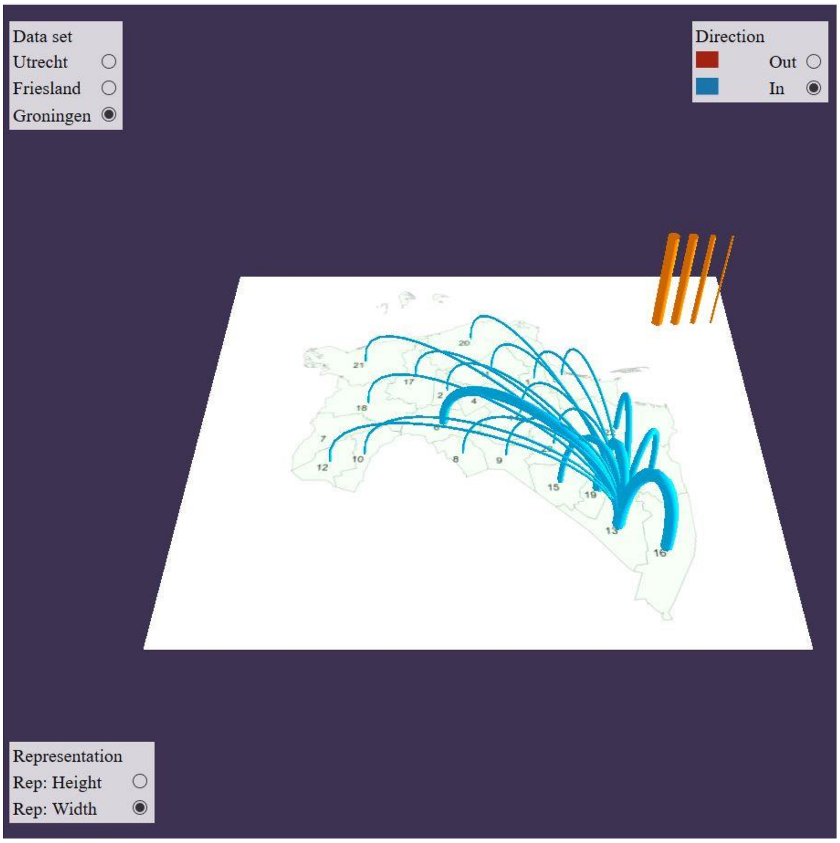
PC 7.2: To which class does the flow from municipality 1-14 belong?

1. *class 1*
2. ***class 2***
3. *class 3*
4. *class 4*

M7: Name the destination municipality with the highest flow departing from municipality 1.

1. [blank] **5**

Map 8.AG4



PC 8.1**:** When comparing the flows between municipality 13-15 and 13-16, which flow is the largest?

1. [*The flow from municipality X to municipality X*] **13-16**

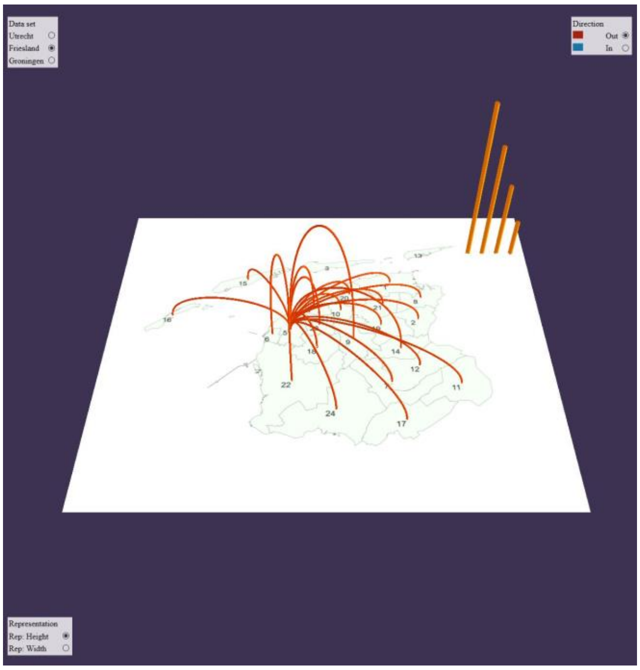
PC 8.2: To which class does the flow from municipality 13 to 22 belong?

1. *class 1*
2. ***class 2***
3. *class 3*
4. *class 4*

M8: Name the origin municipality with the largest flow going to municipality 13.

1. [blank] **16**

Map 9.AF1



PC 9.1**:** When comparing the flows between municipality 5-6 and 5-7, which flow is the smallest?

1. *[The flow from municipality X to municipality X]* **5-7**

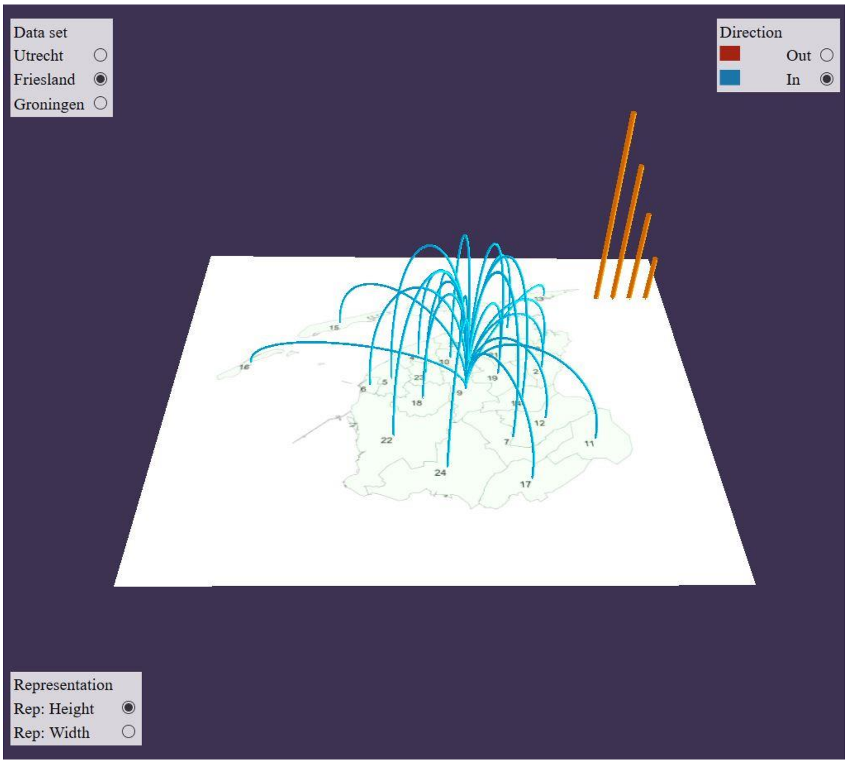
PC 9.2: To which class does the flow from municipality 5-4 belong?

1. *class 1*
2. ***class 2***
3. *class 3*
4. *class 4*

M9: Name the destination municipality with the highest flow departing from municipality 5.

1. [blank] **9**

Map 10.AF2



PC 10.1: When comparing the flows between municipality 9-21 and 9-14, which flow is the smallest?

1. *[The flow from municipality X to municipality X]* **9-21**

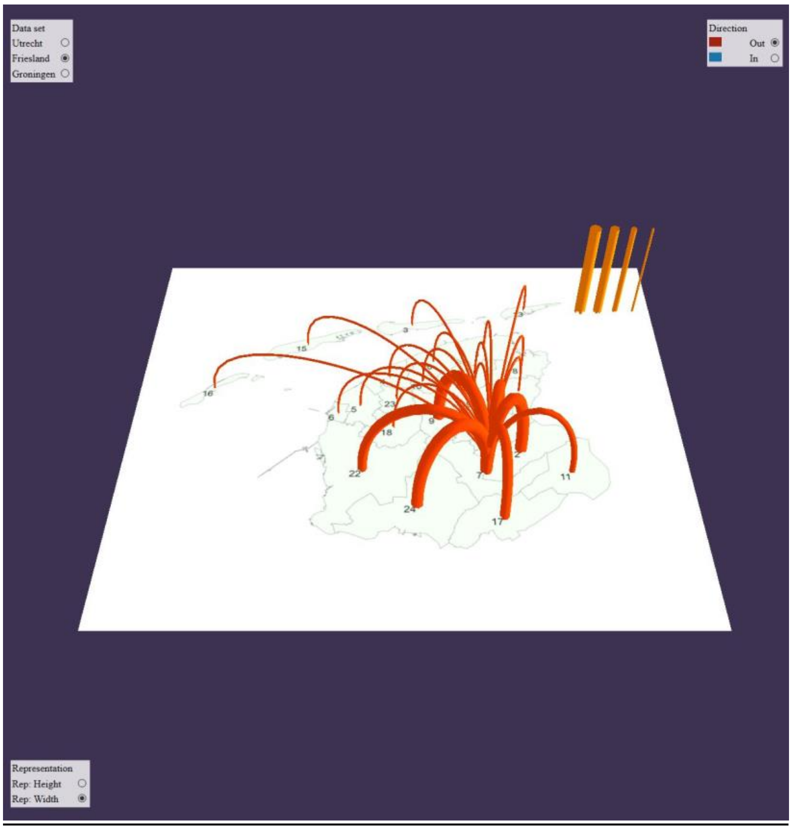
PC 10.1: To which class does the flow from municipality 9 to 2 belong?

1. *class 1*
2. ***class 2***
3. *class 3*
4. *class 4*

M10: Name two origin municipalities of the largest flows going to municipality 9.

1. [blank] **5, 10, 14, 17, 19, 22**

Map 11.AF3



PC 11.1**:** When comparing the flows between municipality 7-11 and 7-12, which flow is the smallest?

1. *[The flow from municipality X to municipality X]* **7-12**

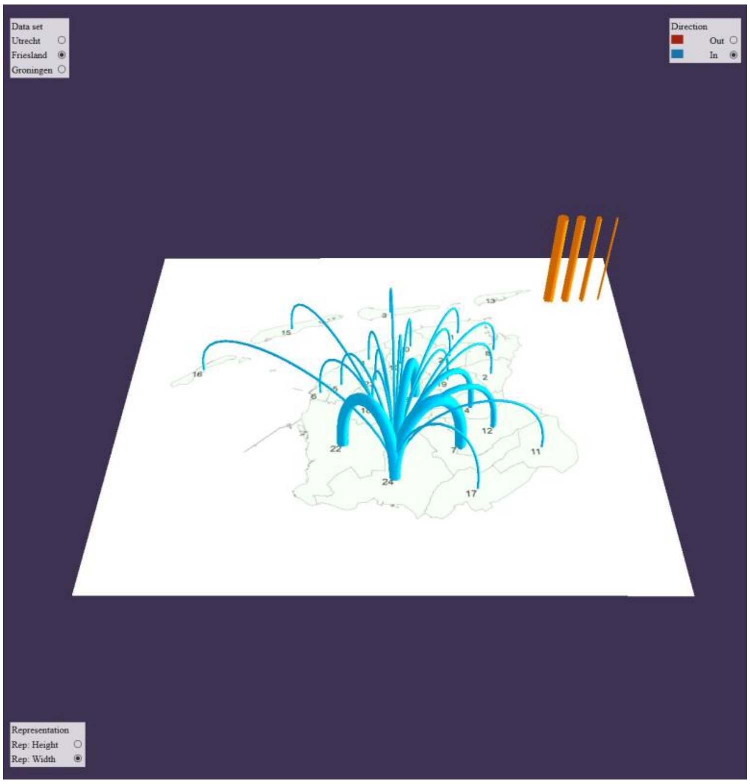
PC 11.2: To which class does the flow from municipality 7-3 belong?

1. ***class 1***
2. *class 2*
3. *class 3*
4. *class 4*

M11: Name one of the destination municipalities with the highest flows departing from municipality 7.

1. [blank] **9, 12, 24**

Map 12.AF4



PC 12.1**:** When comparing the flows between municipality 24-7 and 24-8, which flow is the largest?

1. [*The flow from municipality X to municipality X*] **24-7**

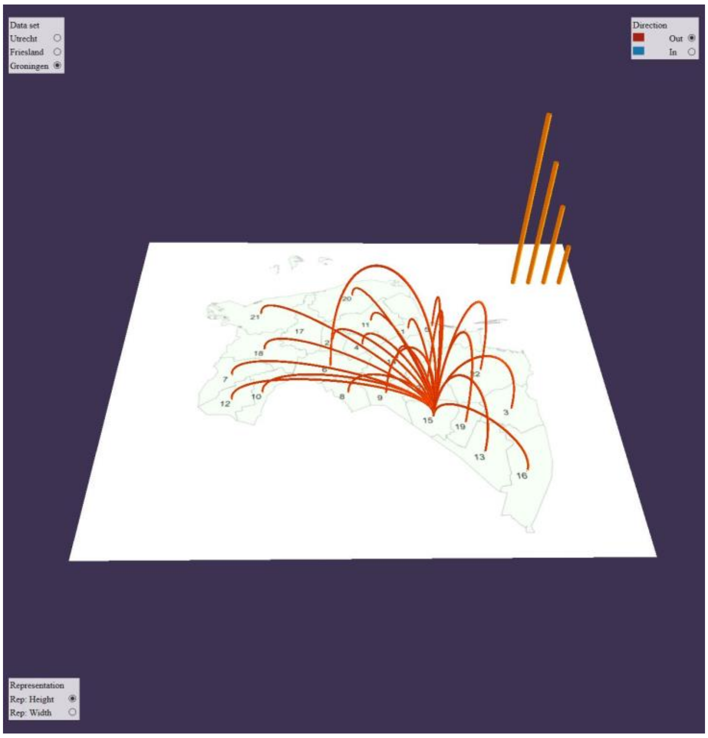
PC 12.2: To which class does the flow from municipality 24-12 belong?

1. *class 1*
2. ***class 2***
3. *class 3*
4. *class 4*

M12: Name the two origin municipalities with the largest flows going to municipality 24.

1. [blank] **7 and 22**

Map 13.AG1



PC 13.1**:** When comparing the flows between municipality 15-9 and 15-10, which flow is the smallest?

1. *[The flow from municipality X to municipality X]* **15-10**

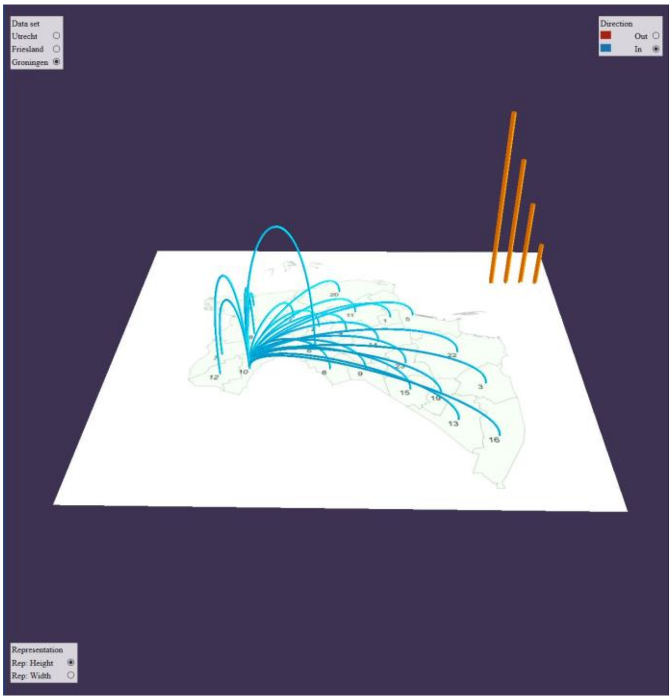
PC 13.2: To which class does the flow from municipality 15-3 belong?

1. *class 1*
2. ***class 2***
3. *class 3*
4. *class 4*

M13: What is the destination municipality with the highest flow departing from municipality 15?

1. [blank] **6**

Map 14.AG2



PC 14.1**:** When comparing the flows between municipality 10-11 and 10-12 , which flow is the largest?

1. [*The flow from municipality X to municipality X*] **10-12**

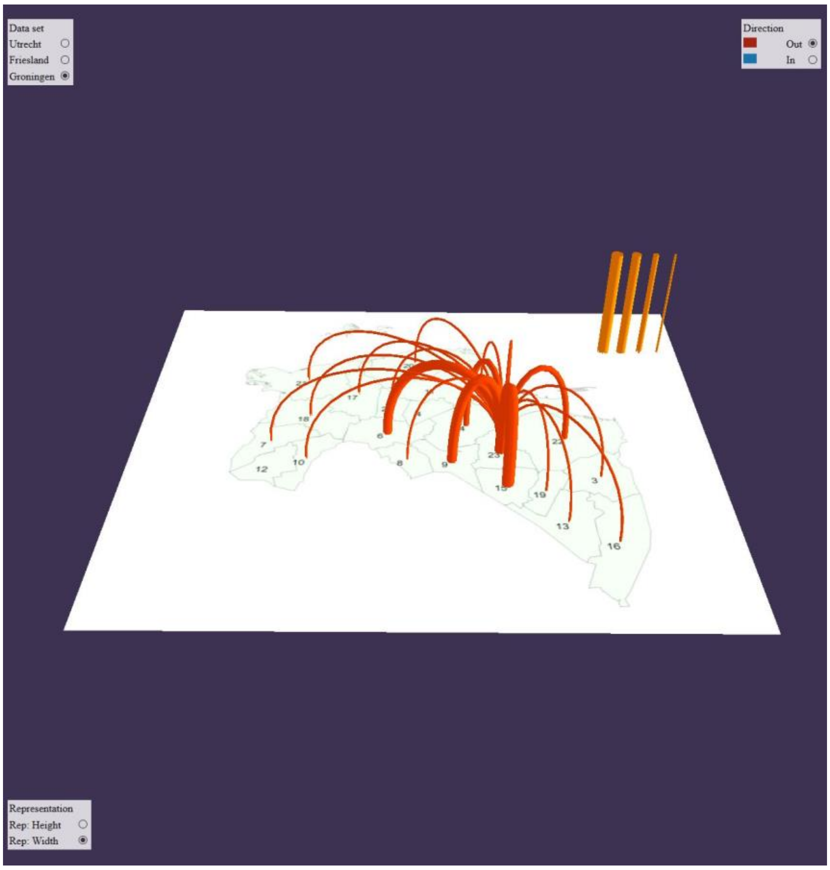
PC 14.2: To which class does the flow from municipality 10-20 belong?

1. ***class 1***
2. *class 2*
3. *class 3*
4. *class 4*

M14: What is the origin municipality with the largest flow going to municipality 10?

1. [blank] **6**

Map 15.AG3



PC 15.1**:** When comparing the flows between municipality 23-8 and 23-9, which flow is the largest?

1. [*The flow from municipality X to municipality X*] **23-9**

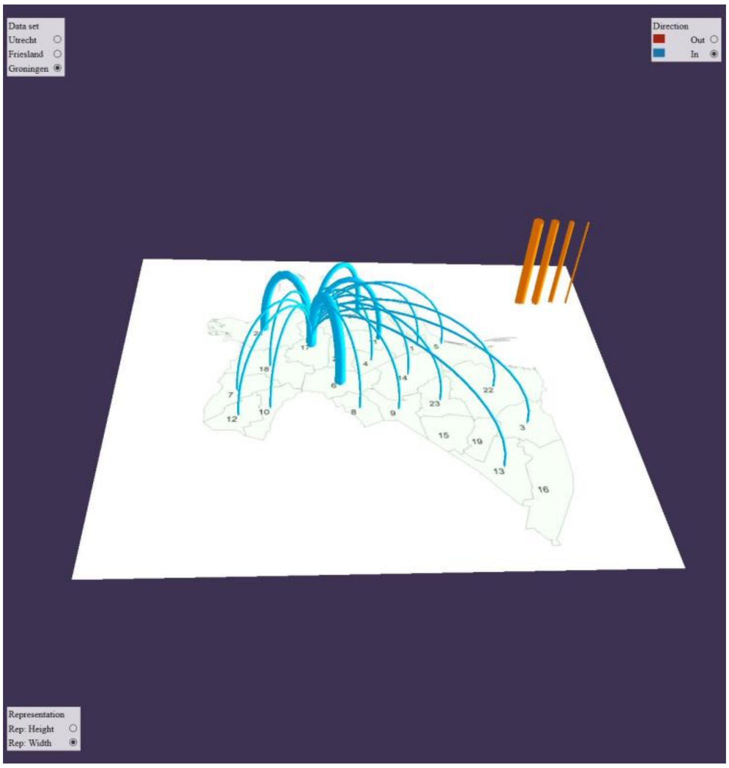
PC 15.2: To which class does the flow from municipality 23-7 belong?

1. ***class 1***
2. *class 2*
3. *class 3*
4. *class 4*

M15: What is the origin municipalities with the largest flow going to municipality 23?

1. [blank] **15**

Map 16.AG4



PC 16.1**:** When comparing the flows between municipality 17-21 and 17-23, which flow is the largest?

1. [*The flow from municipality X to municipality X*] **17-21**

PC 16.2: To which class does the flow from municipality 17-7 belong?

1. ***class 1***
2. *class 2*
3. *class 3*
4. *class 4*

M16 What is the origin municipality with the largest flow going to municipality 17?

1. [blank] **6**

A2. Manual usability test

Instructions

Test overview

The test is divided in three sections:

1. **Introduction (10 minutes)**
   A short survey on personalia and map experience + familiarization with the maps

2. **The usability test (20 minutes)**
   You will be presented with a set of flow maps, representing the migration of people between municipalities in the Netherlands in 2017. These maps are designed in two display environments (on screen and in augmented reality) and the data is represented in four classes with two types of cartographic styles (height and stroke width). You can use the online survey on your smartphone to answer the questions based on the corresponding maps.

3. **Map satisfaction interview (5 minutes)**
   After the test, I will ask you to rate the readability of a selection of the maps.

Measurements

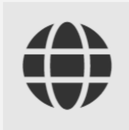
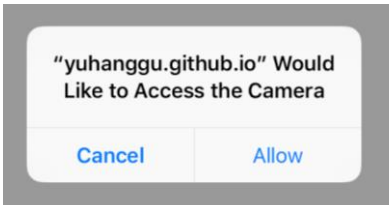
The following information is recorded:

- Your answers
- The time spent for each question
- Your spoken thoughts (by using an audio recorder)
- The screen
- Any additional behaviour or events that could influence the results (by using video camera)

You are assigned to a participant number. All the data will be anonymised, and the audio and screen recordings will only be used for the analysis of the results. The audio and video files will be deleted afterwards. By participating, you agree that this session is being recorded.

Navigation

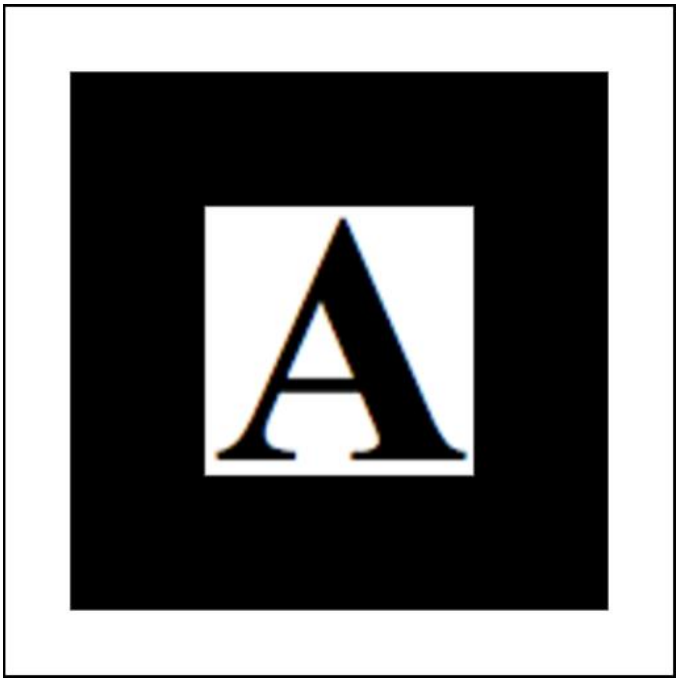
Table 1 - Navigation instructions for the maps in both display environments

	display environment	
tasks	 **PC** monitor	 **AR** iPad
open the map	1. click on the map symbol	1.tap the map symbol 2.allow the following notification: 
return to menu	click on return page (Chrome) 	click on return page (Safari) 
move orientation	mouse left click and drag	move iPad device around the target
pan	right click and drag	
zoom	turn the scroll wheel	

- You can not return to previous questions.
- The test should be performed individually without the assistance of the investigator.
- Please speak clearly, and say anything that you see or think when using the maps.

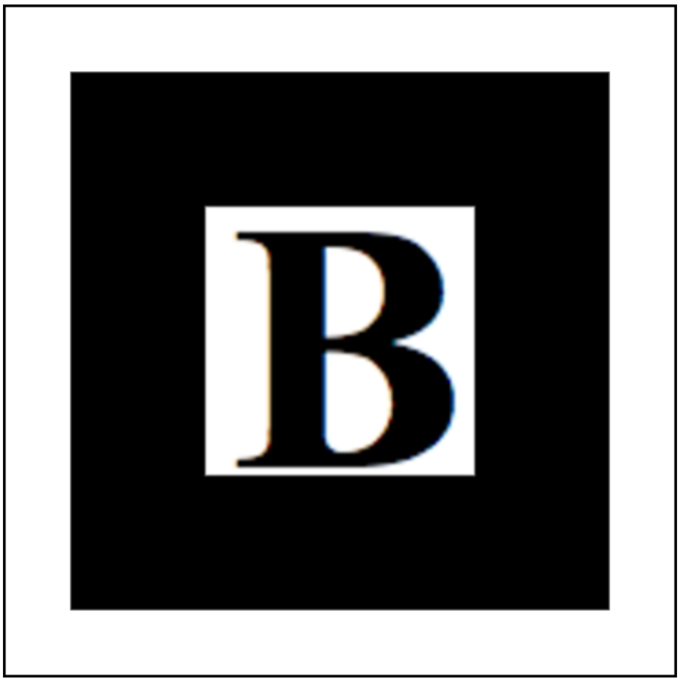
A3. Targets MR

Mark - A



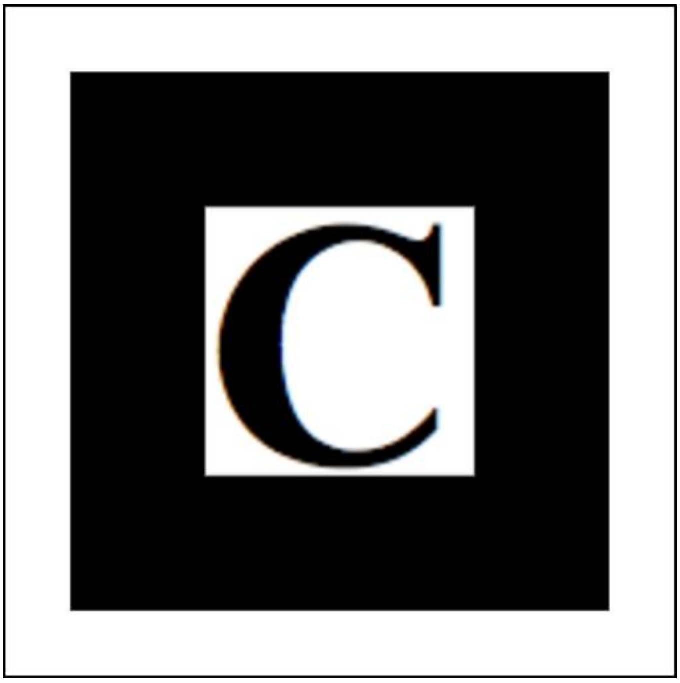
Focus on this letter with camera to get access to map : TU1 / AF1 / AG1 / BF1 / BG1.

Mark - B



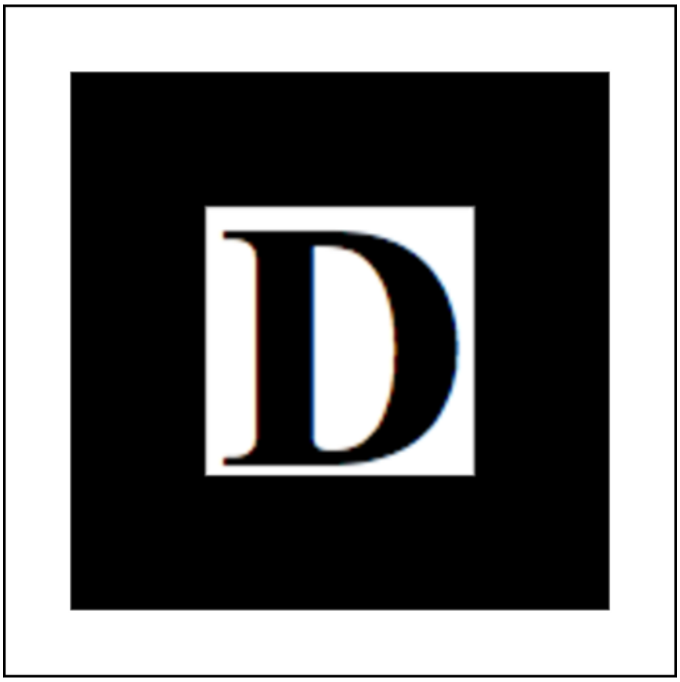
Focus on this letter with camera to get access to map : TU2 / AF2 / AG2 / BF2 / BG2.

Mark - C



Focus on this letter with camera to get access to map : TU3 / AF3 / AG3 / BF3 / BG3.

Mark - D



Focus on this letter with camera to get access to map : TU4 / AF4 / AG4 / BF4 / BG4.

A4. Focus group transcript

Thesis title	The improvement of the visualization of 3D flow maps
Master	Geographical Information Management and Applications, Utrecht University
Student	Niels Struis
Respondents	Corné van Elzakker Ieva Dobraja
Date	June 13th 2019

Transcript of the meeting

Niels
For the methodology, I made a list of all requirements for the test. Let me read it over first.
**see methodology*

This is an overview of all the maps.
**see overview table of the maps*

For the familiarization phase, two maps are used and for the usability test, 16 maps are used. As you can see, all maps are divided into two main groups; A and B to enable a difference in the order of the display environments per participant. For half of the respondents, the maps covered in group A a shown on a computer screen, and the maps in group B on the iPad, while this is switched for the other half of the respondents. The ordering per participant of the display environment can be controlled because of main groups A and B. Subgroups H and W are then designed for cartographic representations of the flow lines. The maps are grouped like this because a certain order might cause biased outcomes. The maps are made by Yuhang Gu.

Corné
Are you working with Yuhang?

Niels
Yes

Corné
Ah. I didn't know that.

Niels
We are collaborating since Yuhang is doing his PhD research on the usability of flow maps. The feedback of this usability test will contribute to his work, while I can make use of his maps.

We used mean values as the classification method to classify the flows per province on the size of the flows. First we used the same classification for all provinces, but this caused many flows belonging to the highest class for Groningen. Now, the flow lines are spread more equally amongst the classes for each dataset.

Corné
I think I need to see the maps to fully understand this. But I have one question beforehand. So you are saying that you are using AR together with the representation on the iPad, and the other representation on the screen. That other screen is larger, probably?

Niels
That is correct.

Corné
And that AR is in itself probably a more cluttered image, because you have the same what you present on the big screen with AR added to it. Wouldn't that influence the outcome? Because you are talking about clutter right? Because of the size I immediately say that is not fair. You should control as many external factors as possible. You are introducing that yourself already, that you want to say the exact same things to the participants because you want to control the experiment. In that sense, I would also recommend to write the instructions down and not explain it verbally because it could be that you will not explain it the same way you you did the first time, as the second, third, fourth and fifth time. Probably, because of your experiences already and that is not fair. They should all be given the same instructions, and give them the same amount of time to experiment. This way, their starting points are controlled as well. Also, the representation medium should be controlled. You should only be looking at the variables; the four stimuli.

Niels
I think it is indeed better to use all maps on the same screen if possible. After Yuhang finished the online application to view the maps, it is now possible to view both maps on the iPad as well. So this problem can be solved. [update: the PC maps are not shown correctly on the iPad, therefore the PC maps will still be shown on a computer screen. The browser window needs to have the same size as the iPad screen.]

Corné
Luckily you have the big iPad. it is already difficult that such type of maps will be consulted on an iPad and not on a desktop screen? For sure, you can assume that this type of maps won't be consulted on a smartphone because that is too small. A big computer screen is even larger. The question is whether you should do it on an iPad or on a larger screen. How many people will actually use it on an iPad in reality?

Niels
I agree with what you said. However, it is difficult to apply AR to a computer screen because you would need a camera. In the case of virtual reality (VR), this could be applied to larger screens or even VR glasses.

Corné
I like to see it.

Niels
This is the application and questionnaire. Here are four maps of the province of Utrecht for which the strokes are represented in height or width. Blue lines indicate all flows going in the municipality, and red going out.

Corné
Ah, so its a map of the province.

Niels
Each participant is assigned to a unique ID, which will have its own ordering of all maps. For the odd numbers, group A is shown on screen and group B in augmented reality. For the even numbers, group A is shown in augmented reality and group B is shown on screen.

This is the legend of the four classes. For the questionnaire, you can put in your user ID first. The order is different for each participant based on these groups. Each participant is going to use all the maps, but in a different order.

Corné
This ordering is very good, because indeed bias can occur.

Ieva
What about the dataset? Do they all have the same number of flows?

Niels
That is a very good question. The province of Utrecht has 26 municipalities, Friesland 24 and Groningen 23 as for the year of 2017. This way, it is quite similar in number especially for the two datasets in the main usability test.

Corné
Who is choosing the maps? Are you as the research leader doing that?

Niels
The user has to do it. Above each question in the mobile questionnaire, the corresponding map number is shown. Hopefully, this is easy enough for the participants but I should always check that the users are using the correct maps for each question.

Corné
In theory it is not so difficult to make the maps appear in a certain order automatically.

Ieva
Will there be a description of the map? Like a title?

Niels
I will discuss this later with Yuhang. As for now, the description of the map is shown in the questionnaire.

Ieva
Also can you click on the municipalities to show the names?

Niels
We chose to use only numbers for two reasons: it makes the geographical area more anonymous and therefore the user is less likely to be familiar with the context of the migration in these areas. Secondly, numbers take up less space which enables a better readability of the flow lines.

Maybe I can show you the questions now. First, the user will fill out some personal questions about their age, gender and study background. After that, the following questions are mentioned to test the experience with flow maps, 3D maps and augmented reality.

Corné
When do you present this? After they have learnt to work with the maps?

Niels
No, this is presented beforehand. The familiarization happens after this screen.

Corné
That is a problem indeed. Do all the participants know what a flow map, 3D map or augmented reality really is? If you ask questions like this, which is very good because you need this information to be able to interpret the outcomes later on. You must either do the familiarization beforehand, or show them pictures of these maps and techniques.

Niels
That is a good idea. So do you suggest to insert these descriptions or images below each question?

Corné
Yes.

Niels
Let's continue with the questionnaire. The following screen shows the start of the familiarization phase along with instructions.

Corné
I would give a little bit more operational instructions instead of you explaining it verbally. I would suggest to make a very short manual. These instructions should be the same for everyone. In research, you should try to prove that you are unbiased. It could be that at a certain moment that the results are such that you don't expect them and are tempted to influence these a little bit. You should give the basic instructions and let the participants just answer the questions without any help.

What you can also do. If you have that short manual on paper next to it on the table as a shortcut.

Niels
So they can look back at it in case they have difficulties with the controls?

Corné
Yes.

Niels
This is a very good point. Next, about the questions. There is three types of questions. The first type is about comparing two flow lines to distinguish which is smaller or larger. The second type is to which class a single flow lines belongs. The third type requires the user to recognise the highest or smallest which is more difficult than the first two questions.
For instance, map TU1 is used for the following questions.

shows the map and goes through the questions together.

Corné
Am I correct that the highest legend bar is not occuring at all?

Niels
I will discuss this with Yuhang as well. It is correct that the height of the legend bar does not correspond with the maps.

Ieva
This is very important.

Niels
I think this is possible for Yuhang to adjust this, but I will ask him later today.

Corné
Like Ieva explained, there is a need for more information on the description. For me it is quite simple, because I have seen this before. But do people immediately understand that the height or width of the flow lines are representing the quantity of the migration? Do you want to find out whether they understand the map that you give them? Or do you want to instruct them?

Niels
It is not the goal to let the users figure out what is what. Everything should be instructed as clearly as possible beforehand.

Corné
So now you talk about height. Maybe it can help if you put below each map somethin like “The height of the curve represents the number of migrants from municipality X to Y.” - or something like that. Just like the direction of the fow.

Likewise, the legend should be more clear. The should be an explanation of the classification below, and then show the classification numbers below the bars.

Niels
Thank you, this is a very helpful remark.

Ieva
And maybe put the legend and map lines in the same color?

Corné
Why are there two colors again?

Niels

First of all, we were thinking of using arrows to represent direction. This was too cluttered and created an extra stimuli which would make the analysis too complicated. Therefore, two colors are used: blue is inflow, and red is outflow.

Corné

That is understandable. So you have volume and height as the two representations?

Niels

Yes. All shown in four classes but in different representations. I will now show the maps in augmented reality. For each dataset, you first have to click on '*allow camera access*' when you open the map. When you move the iPad around the target, the perspective in which the map is shown will change. This makes the navigation different from the PC maps.

Corné

Where is the legend?

Niels

That is a good point. The legend is not displayed on the AR maps yet, but this will be integrated soon.

Corné

About the terminology, I would call this mixed reality instead of AR. This you write about this, Ieva?

Ieva

I read about it. For me it was not really clear what mixed reality is. I can't really distinguish it from AR.

Corné

I supervised a student who made a very nice graph with the two different ends; real reality and virtual reality. Pure AR in my view is; you look at reality. For instance, a ar mechanic who looks at the motor, and gets information through labels of what he sees there (oil thank, etc). IN map applications you see that very often for tourist information when you stand on a square, you point your camera around you and you get information about a sight. That is augmented reality. You augment the reality with additional information. Here we are not working with reality. We a working with a display on a table which is nice. Is it reality? You focus on the interface that is different. I my view, the main differences between your display environments are about the interaction/navigation with the data. Which is fine, but it is a matter of terminology.

Niels

I agree. For my thesis, we used AR as the technology to display the map, which is an enclosed way of using this technique. For instance, for Google Maps you can see the arrows on the street, in the real world around you while this is based on only one target.

Corné

I am not questioning it. It is just a matter of terminology, I got the wrong impression beforehand. What I will do, Xiao Linn did augmented reality. She went out in the field with her application and derived information about textile factories. That was pure augmented reality. I can send you a thesis, I think it was a MSc. cartography student who used a hololens in Munich. There we also had the discussion on the terminology of VR, AR and MR.

Niels
I shall critically look at this terminology again.

Corné
The reality you are talking about is arbitrary, it doesn't mean much, as a piece of paper on a table. In AR, the starting point is reality, you simply add something to it to make it understandable. The same goes for the kind of 3D that we are talking about. What is represented? Maybe it is 3D on a 2D screen? It is often called pseudo 3D or 2.5D. This is not essential for the usability test, because they won't care and it won't influence the results. However it is essential for the thesis reporting.

Niels
Thank you! About the AR maps. All maps are similar in nature, but with a different dataset, cartographic representation, direction and origin/destination municipality.

**explains the AR marks*
The rest of the survey goes with these three type of questions, for all 16 maps. I was thinking, 4 maps might be too much for the familiarization.

Corné
I think so too. In any case, you should do a pilot test, after which you can make adjustments before you start with the actual test. This is important because you do not want to spoil any potential participants. This will help you to save time, if you are testing what you want to test.

About the ordering of the maps. It is good to show all stimuli to all participants. The alternative is to make groups of the participants, but this makes it difficult to compare the stimuli. Your participants are of a specific study background. It is not bad to have people who are at least a little bit engaged with these types of maps. I think the way you have grouped the maps is good.

Niels
I have some additional questions. In total there are 16 maps with 3 questions each. The questions are fairly easy, and therefore I assume that this would not be too long. What do you think?

I was not sure if the question types are challenging enough. For instance, Yuhang proposed another type; "How many flows belong to this class?"

Corné
The questions should be realistic, not too technical. Participants should get a good understanding about the migration of the people moving between the municipalities. An example is "What is the trend of the flows in this map?".

**Corné brings a book*
These are different types of difficulty level questions. I think it is good to have these three levels of questions in your task. Elementary, comparison and overall. These examples are for another purpose, not for flow maps, but it can help as an example. You can say something like, what is the trend? With three alternative closed answer possibilities.

Niels
For the next question, which device would you think is most suitable to et the participants fill out the answers? A laptop or a smartphone?

Corné
I would say paper.

Ieva
I would say laptop

Niels
The advantage with a smartphone / laptop is that you can record the time.

Corné
I see. But you don’t want to record the sessions?

Niels
Yes, as a backup or in the case an external event might influence the results.

Corné
That is very good. The big advantage of recording through video is that you can interpret the behaviour of the participants a little better. ideos give insight in the engagement of participants. With audio, you might get the wrong impression about their enthusiasm, for instance. The think aloud and video are a good backup for this. So imagine, when someone starts talking about soccer for instance, that would not be interpreted when the time is only recorded by the video/audio. It is important that you mention to the participants that their privacy is maintained.

Niels
Thank you. Lastly, about the third component; the satisfaction interview. For a selection of 8 maps, with an equal variety of the stimuli, the user is asked to assign a Likert scale score to the readability.

Corné
It is good that you are asking the satisfaction afterwards, including the thumbnails of the corresponding maps. Or they can go back to the previous maps to check it again. If you do not mind I have to leave. You can have the book.

Niels
This was all I wanted to ask. Thank you very much for participating!

A5. Notes of all sessions

For the participants who talked in Dutch, their comments are translated to English.
An estimation of the **delay** is given in bold red text (exact times should be traced back by using the recordings).

All participants are asked to put their phone notifications to silent.
Having sufficient light in the room is a requirement for the AR maps.
Niels checked whether the participants were using the correct maps, which means that he had to indicate the right AR targets for some participants.

Some participants talked more than others. Niels reminded the people who remained silent multiple times to speak out any thoughts, observations and reasonings for answering the questions, nut only in some instances did these efforts made by the investigator results in additional comments by the participant. Some people might not find it logical to speak out their thoughts while performing the test, or were too busy figuring out the answers to the questions that they kept forgetting to add their thoughts. People who speak more, often times give critical comments as well. It should be considered that participants who speak more, are likely to need more time to finish the test.

The comments from the satisfaction interview are not included.
The screen recording errors are shown in table 1.

Table 1 - Overview of all collected recordings per participant ID. The blue cells indicate the missing recordings.

ID	iPad screen	PC screen	video	audio
1	X	✓	✓	✓
2	✓	✓	✓	✓
3	✓	✓	✓	✓
4	✓	X	✓	✓
5	✓	✓	✓	✓
6	✓	✓	✓	✓
7	✓	✓	✓	✓
8	✓	X	✓	✓
9	✓	✓	✓	✓
10	✓	X	✓	✓
11	✓	✓	✓	✓
12	✓	✓	✓	✓
13	X	✓	✓	✓
14	✓	✓	✓	✓
15	✓	X	✓	✓
16	✓	✓	✓	✓
17	X	✓	✓	✓
18	✓	✓	✓	✓
28	✓	✓	✓	✓
29	✓	✓	✓	✓
30	X	✓	✓	✓

ID 1

18-06-2019 11:30

AF1 mistake in the question; not from 19 to 7 but from 9 to 7. [This has been fixed after this test] **delay 10 sec**

AF3 1: “Is it correct that multiple flows can belong to the largest class?” Niels: “Yes”.

AG1 “There should be a button that takes you back to the first map view, to reorient the map to the central position”

-- switch to AR --

“The targets for the AR maps are being recognised by the iPad very quickly”

“Municipality 6 is often times the answer for the largest flow(s) in the Groningen. At a certain moment, you can familiar with this”. *opens next question where another flow os the largest* “Or maybe not?”

ID2

18-06-2019 13:00

AF1 Started with the wrong map, screen recording is restarted. **delay 2 minutes**

“It is difficult to trace individual lines.”

The participant holds the iPad vertically in order to see the heights of the lines more clearly. “Because of the perspective, it is difficult to compare the lines with the height of the legend.”

Seems to take more time for the questions than the previous participant.

“it’s difficult that all lines have the same shade of blue.”

ERROR: The questionnaire page was closed, so he started over by filling in his ID but random answers up until where he left to map AF2. Therefore, the data from AF1 was not saved. **delay 1 minute**

People seem to become quicker at responding to the questions as they progress with the test. This can be a prove of the familiarization effect of both the data as well as the controls. (The 2 maps for the familiarization might have not be enough to fully reach everyone’s full potential, however extending this phase would extend the total test duration which is less realistic to perform and the focus of the participants might reduced)

Participant is actively moving the iPad to view the maps.

AG3 seemed to be easy (very quick responses).

-- switch to PC--

“There should be a button to go back to the central view”

There were barely any uncertainties about the navigation controls of the PC map.

Faster navigation on the PC than with the iPad. Using the mobile questionnaire in combination with the PC maps seems to be an easier operation than in combination with the iPad earlier.

AF1“This is difficult because many flows are clustered together because these belong all to the same class”

ID 3
18-06-2019 14:00

AF3 “It is not clear which line is larger.”

AF4 “I am doubting when a line belongs to a municipality”. (municipality 7 for the Friesland maps)

It becomes clear that using a smaller browser window than usual (for having an equal display size as the iPad), it is more difficult to have the legend visible along with the rest of the data. This might require more acts to get both the map and the legend in the view again (panning/zooming)

AG1 Niels: Multiple answers are possible for question 3, but please only mention 2.

-- switch to AR --

The user is holding the phone and iPad simultaneously.

“Again municipality 6.”

Satisfaction comments
AG3 “For the stroke width, the thickest and thinnest lines are easy to recognise, but class 2 and 3 are difficult to distinguish.”

“A center view button would have been nice for the PC”

ID 4
18-06-2019 16:00

Niels had difficulty with arranging the next room, which forced him to wait an hour before the next test could begin.

Niels had a hard time to see when the participant was proceeding to the first question of the test, because you cannot always look on their small phone screen. This is why he always explained to tell him when they have answered all the questions for the AR Utrecht map, so all recordings can be turned on before the first question is opened. However, this did not always work out well. This is why the weighting of the first map should be smaller, because it might vary more amongst the participants because of these delays.
“You do not have to move the iPad as much for the maps showing thickness than for the maps showing heights”

“I am unsure if I accidentally answered the smallest/largest municipality for the previous question. → This indicates that reading the questions can be a challenge and therefore they should be focused at all times to find the correct answers from the map.

He goes through the maps quickly. However, this speed of working might vary amongst participants. Where one is more precise before giving their answer, the other seems to base its answer on the quick first sight.

“It becomes more difficult to read when all flows are coming together to the center of the map.”

“The maps were easier on the PC, but the navigation with the mouse orientation was not ideal.”
“Stroke width is better for comparing between the flows.”

“In some instances, I had the feeling that the largest class did not occur in the height maps, but this might be because of the perspective effect when the legend is located further away in the background.”

ID 5
18-06-2019 20:00

Participant started with the wrong map; recordings are restarted except for the questionnaire. **delay 1,5 minutes**

“I tend to look more at the differences between the flows than to use the legend”

AG3 Map did not load immediately, so had to refresh the page. **45 seconds delay**
AG4 “One class is missing?”
ERROR: 9 to 8

-- switch to AR --

“This is more difficult than on the PC”
She hold the iPad close to the targets, maybe because she needs to zoom in for comparing individual lines.

‘It is more difficult to zoom in on the iPad, because then you get too close to the target”

ID 6
19-06 09:00

“It is like a puzzle, not quite a user-friendly map.”

“Why don’t you use different colors per class instead of heights/widths?”

“Municipality 7 is at the border with 12. As a GIS student, I see that you have applied centroid for placing the points, but this can be unclear for many people.”

“I prefer the blue lines above the red lines”

“it is annoying that the legend is not always simultaneously on the screen with the rest of the map. As soon as you look at the map from the right angle to observe a specific line, the legend might not be easily in reach to compare it with.”

“For some maps the questions are very easy, for others this can be unclear.”

AG4 Niels had to explain the mistake in the question **delay 10 sec**

“I would have preferred to be able to hold the iPad at chest level, because that would make it easier to watch.”

“What does the blue/red lines mean again?” Niels explains. “I would have added a small arrow going to the left or right at the bottom of the legend foreach map. I am just not good with colors.”

ERROR: AR map is flickering.

“The height maps are more clear, however for these I need to zoom out more.

AG2 Niels explained that more answers are possible **delay 10 sec**

‘After a while, you get used to the maps such as bringing the legend in the view more clearly.”

-- switch to PC --

“Sometimes, there are that many lines so that I have more difficulty reading the municipality numbers underneath.”

Legend class is put on top of number 13

“I get annoyed when I lost control over the navigation (when map turns upside down for instance). The navigation of the computer is more comfortable, however this is not optimal. The legend is more clearly visible on the PC.”
“I keep forgetting what the direction of the blue or red colors is.”

“The nice thing about the iPad is that you can hold is vertically, and see the height differences in one instance from a side view.”

“These maps are unclear for the average layman”

“If you want to show different migration flows at the same time, I could also use a query in ArcGIS or a complete table with a map with only flows of the outliers. Why are the maps not interactive so I can make highlights to the lines?”

ID 7
19-06-2019 10:30

it is difficult to see the difference in width. Difficulty with holding the map controls straight without moving the map upside down.

He was quickly familiar with both the PC/AR controls.

AF1 “For this map, I need to compare the length now.” Niels: “for these maps, the difference in height is used” (municipality 13).

“I prefer the blue lines because it is easier on the eyes”
“When you get the hang of it, it goes faster.”

-- switch to AR --

AF4 map disappeared for a few seconds - all he had to do was to re-aim the iPad at the target.

The maps with height are more difficult to read.

General remark: He did not speak much, this might be one of the reasons why is was faster.

“Number 8 fell outside the map polygon, on the black surface of the target which makes that number more difficult to read.”
“It is nice to be able to view the maps from above for comparing the widths.”

ID 8
19-08-2019 11:00

AF3 the user started with the wrong map **delay 15 secs**
“It is hard to handle both the iPad and phone at the same time” During the test, he switch the two devices between the hands.

AF4 Sometimes the maps are flickering in AR.

The height maps are working better, and are not as crowded to the ‘focus municipality’.

AG4 switched to the wrong map **15 sec delay**
AG1 Niels explained how multiple answers are possible; please name two. **delay 10 seconds**

-- switch to PC --

Navigation is easier because I am much more used to using a mouse.
BC3 switched to the wrong map **10 sec delay**

Height is way better than the width for AR. PC is much easier to navigate. When the location of the focus of the flows is in the center of the province, it makes it more readable.

ID 9
19-06-2019 13:00

Participant discovered a new method for zooming in on PC map during the familiarization Cntrl + scroll: zoom in on the display view. I allowed this.

AG1 Niels explained how multiple answers are possible; please name two. **delay 20 seconds**
Took a long time for the first map, it seems like he looked in much detail before giving answers.

AG2 “Why is the legend the same color as the map?” Niels explained **delay 10 seconds**
BF4 “What does the red/blue lines mean?” Niels explained **delay 1 minute**

Already clicked to next question before switching between the environments **delay 1 minute**
BG3 number 8 falls outside of the map.

He did not remember the difference between the red and blue lines.

General remark: participant was very silent, but one time interrupted the test for a large comment. (BG3)

ID 10
19-06-2019 14:00

For answering the questions, I first look at the map from above to find the municipality number, then I look at the flow and change the angle to view the map from the side to see the differences in height.

AG1 Niels explained how multiple answers are possible; please name two. **delay 10 seconds**

“It is nice to be able to look at the width map from above to quickly see the differences in flow sizes.”

“The iPad is heavy. Maybe this is a good sign, because it means that we are not yet used to holding a tablet in our hands all the time”

“I find it difficult to see where the lines are going to, but the blue lines are easier to read.”

“It is easier when two municipalities that are being asked are located close to each other”

“Stroke width maps are easier to read”
-- switch to PC --

PC questions are answered more quickly.
“it is nice that there is a uniformly colored background for the PC maps”

“AR maps are very exciting, but it does not add value to the communication of the information in my opinion. I think it can be an effective technique for maps like this, but in this instance it is not very useful yet. Who knows, maybe in the near future we can use this technique to quickly get additional information on nearby objects or buildings, I am thinking of futuristic devices such as Google Glass.”

ID 11
19-06-2019 15:00

participant’s phone battery was low, so the investigator’s laptop was used to fill out the questionnaire. (this was an improvised solution but it should be considered that the answering time could be quicker because the participant was able to type it out on a laptop instead on a phone).

The screen recording of the first map was missed.

-- switch to AR --

It seems that the control of the questionnaire laptop is quicker.

At 15:57 a professor walked in the room, because there was a miscommunication of which room was booked. However, the participant continued working on the test but there might be a delay. This has to be checked on the video recording.

BF3 ERROR: switching maps for the new dataset, used the wrong map. Ignore the data for this map

BF4 **delay 1 minute** because of the previous error

BF2 During the last map the video recorder stopped on its own. The camera stops every time after 30 minutes, which is only apparent when a participant takes more time.

“The question ‘Which flow is larger?’ appeared more often, than for ‘smaller’. This could result in the participant automatically assuming that the next question is about ‘larger’ as well, by misreading the question. It is just a piece of critical advice for the questionnaire.”

“You have to perform more operational steps with the iPad because you have to view from any angle, and sometimes stand up or sit down. This makes it take more more time than the PC maps. I suppose you want to have as much clarity as possible for any map?”

“The stroke width maps can be read within a glance from above, while the height maps require you to move in multiple angles.”

ID 12
20-06-2019 09:30

AG1 Niels explained how multiple answers are possible; please name two. **delay 10 seconds**

participant seems to be quickly familiar with the controls.

“The difference in stroke width is not very clear”

“The maps with the stroke width are more clear than for the height, but the difference between class 3 and 4 is unclear. When there are too many lines in class 4 for a height map, this becomes more difficult to read.”

General remark: participant did not say much.

ID 13
20-06-2019 10:50

ERROR: participant started with ID 12 instead of 13, but started again with the correct ID number.

The speed seems to increase as the participant is progressing with the test. Had difficulty with the controls of the PC.

General remark: participant did not say much.

ID 14
20-06-2019 11:35

“AG4 is very clear”
The participant was quick with getting used to the AR navigation.

Yesterday’s bug fix was applied to the wrong map..

ID 15 (only a part of the test seems to be on the server, however these are the notes)
20-**06-2019 14:50**

AG3 clicked on the wrong map delay 1 minute
Goes quickly through maps

“It is hard to navigate the PC maps. You expect it to move straight but you have to bend it and improvise.”

Seperating class 2 and 3 is very hard for the width maps.

AF1 user lost control

“I know the answer, I just don’t know the number, it’s blocked by the legend.” Niels: “That is number 13”.

-- switch to AR --

“The height is way easier for AR. Especially for the question which one is larger?”

I let the participant move the paper target by hands because we were in a small room where is was difficult to stand up and move around. This would not influence any of the results, because the participants who did this only figured this out themselves, Niels never gave them advice to do so.

“This is a heavy iPad.”

Switch between dataset resulted in wrong map delay 1 minute

ID 16
20-06-2019 15:50
Took more time for familiarization

“Sometimes it is more difficult to read the numbered labels, but then you have to do some extra effort by moving around. Not really a problem.”

AG4 “For me, it is more difficult to read the questions properly than the maps, because especially small differences in smaller/larger, etc” **delay 20 seconds**

He moves the target.

“I find the stroke width maps better to understand because the depth effect makes it difficult to differentiate between the classes and compare the flows’ heights with the legend.

-- switch to PC --

BG3 **10 sec delay**
“The legend is often not visible. Additionally it is impossible to return to the first center view. You need the legend to differentiate amongst the individual class numbers, however for relative differences between the flows you can read the map without the legend. **delay 30 sec**”

BG4 “I am annoyed that I do not know what exact range of absolute values each class stands for.” **delay 10 seconds**

BG1 “Not clear to me which flow belongs to which class. Maybe if the label numbers could move along in 3D when moving the iPad? Although this could make the map too crowded. Now I realise that I might have answered the previous question incorrectly.” **delay 20 sec**

Sometimes it is difficult to see which is the largest, purely because of the navigation and perspective distortion.

BF4 line is on the border with mun 12

Satisfaction feedback
Uncertainty can also be based on the nature of the data. He prefered to move the target because this makes it easier than only be able to move the iPad.

ID 28
20-06-2019 21:00

ERROR: iPad had to adapt to the brightness

“I have to get on top of the map to discover the number, and then move in front to see the height.”

He is moving the targets around.

AF3 “When comparing flows, the width maps are easier but if you have to tell which is larger, I am not so sure. Even if I change perspective, I am not sure which is bigger.”

AF4 This is easier
AF1 This is a messy one. I have to change a lot of perspectives but I know it’s worth it because I can always see the difference and I know I’m right.

-- switch to PC --

“The controls are very sensitive, I wish it wouldn’t turn all the way upside down.”

BG3 **20 sec delay**
BF3 was not loading, so the page was refreshed. **10 sec delay**

Satisfaction feedback
AF1 has a little bit of visual clutter, because the lines all come down to the middle
AF3 Easier to read, but I do not like it.
AG3 Don't like it so much
BF1 So far my favourite
BF3 What is class 2-3
Seems to be quick with adapting both AR and PC controls.

ID 29
21-06-2019 09:45

The participant moved the targets.
AG3 did not load, so the page was refreshed **10 sec delay**
"I can hardly see the difference between class 2 and 3 for the width map."

Sara controlled the map more calmly with the mouse, right from the start. This made the maps turn upside down fewer times. This is an example of user behaviour of the maps/hardware.

AF1 Niels explained that the hidden number behind the legend is 13. AR brightness error **delay 10 sec**

BG3 Niels explained that the less clear number is 8
BG1 map flickering; because the held the iPad too close to the target.

Screen recording just started at BG3.
The participant tried to use portrait mode, but Niels did not allow this. **5 sec delay**

BF2 explanation that the line belongs to mun 7
"Maps with stroke width are more difficult to understand"

"It is easier when the lines are clustered at the edge of a province"

General remark: the participant remained silent.

ID 30
21-06-2019 12:23

Explain how to answer to the third question with a comma **delay 20 sec**

"You have to reorient how you look at the map when coming from width to height, because you can come closer to the width map but need to zoom out for the height map."

General remark: the participant did not speak much.

ID 17

22-06-2019 12:00

AF1 clicked on the first map before starting with the test **delay 10 sec**
First min is not recorded on video because of memory card error.

AF2 "I think the highest class is not covered. Wouldn't it be more logic if the numbered labels are assigned to the municipalities in a more logic order so you don't have to search as much as now?
The participant goes through the maps rather quickly.

Every time it is a matter of switching angles, when you have a map from width to height.

BF2 **delay 5 sec**

The height is easier to read on the iPad, but the the general iPad navigation is annoying. In general, the map reading is clear. But sometimes you need to perform some extra operations to find your answer.

ID 18
22-06-2019

As this has no background in geography or GIS and is the only one who belongs to the older age category, the results might differ from the rest of the sample in terms of efficiency and/or effectivity.

The participant does not move the iPad much at the beginning. She had to stand up and sit down to answer each question.

ERROR: it was difficult to load the AR map when the iPad causes a harsh shadow on the target

AG1 Phone had to be unlocked **delay 10 sec**
AG2 got distracted by the other maps **delay 10 sec**
AG3 the participant barely has the legend in view, and probably only measures the relative differences between the lines
AG4 First looks at the map from above, then from the front/side view

-- switch to PC --

Seems to be faster on the PC.
BF3 lost control over the navigation, the map turned upside down. **delay 10 sec**

Screen recording PC error: the disk space was full; the recording could not proceed.

Satisfaction
The difference in height requires more effort to figure out the answers, while the differences in the maps with stroke width can be seen in a glance.
AG1 is more clear because less lines are shown.

I prefered using the computer, because I had more control over the navigation than with the iPad.

General remark: the respondent did not speak much